

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:26:27 ; Search time 50 Seconds  
(without alignments)  
8773.789 Million cell updates/sec

Title: US-09-126-816b-6\_COPY\_1\_1700

Perfect score: 8677

Sequence: 1 MNLVYKALQKQVYVYKRIQ.....YLYGIDRYVNVKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_page.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8646	99.6	2364	2 Q46342	Q46342 clostridium
2	6761	77.9	2367	2 Q9EXR0	Q9EXR0 clostridium
3	6732	77.6	2367	2 Q9F931	Q9F931 clostridium
4	6711	77.3	2367	2 Q46034	Q46034 clostridium
5	2144.5	24.7	554	2 Q93L39	Q93L39 clostridium
6	2070.5	23.9	2178	2 Q46149	Q46149 clostridium
7	1911	22.0	698	2 Q68653	Q68653 clostridium
8	1893	21.8	698	2 Q9EXQ8	Q9EXQ8 clostridium
9	1887	21.7	697	2 Q86141	Q86141 clostridium
10	1469	16.9	553	2 Q9FCX5	Q9FCX5 clostridium
11	501.5	5.8	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	497	5.7	2792	5 Q814R2	Q814R2 plasmodium
13	496.5	5.7	3169	2 Q82916	Q82916 escherichia
14	494	5.7	3317	16 Q8EWP8	Q8EWP8 mycoplasma
15	478	5.5	2771	5 Q26216	Q26216 plasmodium
16	463	5.3	2752	5 Q9BJY0	Q9BJY0 plasmodium

17	460	5.3	5767	5 Q81525	Q81525 plasmodium
18	446	5.1	3130	5 Q9BK46	Q9BK46 plasmodium
19	446	5.1	3130	5 Q81DX6	Q81DX6 plasmodium
20	446	5.1	3254	5 Q9BK45	Q9BK45 plasmodium
21	439.5	5.1	2747	5 Q9BJX9	Q9BJX9 plasmodium
22	437	5.0	2269	5 Q26223	Q26223 plasmodium
23	436.5	5.0	2849	5 Q81HY4	Q81HY4 plasmodium
24	436	5.0	3381	5 Q812V4	Q812V4 plasmodium
25	427	4.9	2184	5 Q81LO6	Q81LO6 plasmodium
26	424	4.9	2166	16 Q51465	Q51465 borrelia bu
27	423.5	4.9	4688	16 Q9PQ08	Q9PQ08 ureaplasma
28	410.5	4.7	6761	5 Q81C77	Q81C77 plasmodium
29	406.5	4.7	1711	5 Q8MWP2	Q8MWP2 plasmodium
30	406.5	4.7	1956	5 Q81IE1	Q81IE1 plasmodium
31	404.5	4.7	1713	5 Q8MWP1	Q8MWP1 plasmodium
32	404.5	4.7	3597	5 Q81LR5	Q81LR5 plasmodium
33	403	4.6	1686	5 Q81FP9	Q81FP9 plasmodium
34	403	4.6	3225	16 Q9PKM6	Q9PKM6 chlamydia m
35	399.5	4.6	2511	5 Q81L44	Q81L44 plasmodium
36	399.5	4.6	3322	5 Q81KLO	Q81KLO plasmodium
37	398	4.6	1979	5 Q96133	Q96133 plasmodium
38	397	4.6	5687	5 Q81JH4	Q81JH4 plasmodium
39	396.5	4.6	3223	2 Q81LL9	Q81LL9 escherichia
40	396	4.6	1697	5 Q81FM4	Q81FM4 plasmodium
41	396	4.6	1716	5 Q8MWH2	Q8MWH2 plasmodium
42	394.5	4.5	3223	2 Q9RPH1	Q9RPH1 escherichia
43	394.5	4.5	6473	5 Q81KH9	Q81KH9 plasmodium
44	394	4.5	2867	5 Q9N2M3	Q9N2M3 plasmodium
45	394	4.5	2910	5 Q81BY8	Q81BY8 plasmodium

## ALIGNMENTS

RESULT 1

Q46342	ID	Q46342	PRELIMINARY;	PRT;	2364 AA.
AC	Q46342;				
DT	01-NOV-1996	(TRENBLrel. 01, Created)			
DT	01-NOV-1996	(TRENBLrel. 01, Last sequence update)			
DT	01-MAR-2003	(TRENBLrel. 23, Last annotation update)			
DE	Cytotoxin L.				
OS	Clostridium sordellii.				
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;				
OC	Clostridium.				
OX	NCBI_TaxID=1505;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=6018;				
RX	MEDLINE=95369733; PubMed=7642137;				
RA	Green G.A., Schue V., Monteil H.;				
RT	"Cloning and characterization of the cytotoxin L-encoding gene of				
RT	Clostridium sordellii: homology with Clostridium difficile cytotoxin				
RT	B.;"				
RL	Gene 161:57-61(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=6018;				
RX	MEDLINE=96149194; PubMed=8544213;				
RA	Green G.A., Schue V., Girardot R., Monteil H.;				
RT	"Characterisation of an enterotoxin-negative, cytotoxin-positive				
RT	strain of Clostridium sordellii.;"				
RL	J. Med. Microbiol. 44:60-64(1996).				
DR	EMBL; X82638; CAA57959.1; -				
DR	InterPro; IPR002479; CW binding.				
DR	InterPro; IPR001950; TIF SU11.				
DR	Pfam; PF01473; CW binding 1; 18.				
DR	PROSITE; PS01118; SU11.1; 1.				
SQ	SEQUENCE 2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;				

Query Match 99.6%; Score 8646; DB 2; Length 2364;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1696; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 MNLVNAQIQKQVYKFRIOEDYVAILNALBEYHNMSSESVVEKYKLKDINNTDNYL 60
Db 1 MNLVNAQIQKQVYKFRIOEDYVAILNALBEYHNMSSESVVEKYKLKDINNTDNYL 60
QY 61 NTKKSGRNKALKKPKXYLTMEVLEIKNNSLTPVEKNLHFIMIGQOINDTAINYNQWD 120
Db 61 NTKKSGRNKALKKPKXYLTMEVLEIKNNSLTPVEKNLHFIMIGQOINDTAINYNQWD 120
QY 121 VNSDYTVKVFYDSNAPLINTLTKTIVESATNTLSEFRENLDNDFDYNNKFRKRMETIY 180
Db 121 VNSDYTVKVFYDSNAPLINTLTKTIVESATNTLSEFRENLDNDFDYNNKFRKRMETIY 180
QY 181 DKQKHFIDYKSOIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLUNKITANNNDI 240
Db 181 DKQKHFIDYKSOIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLUNKITANNNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDLFKS 300
Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQDLFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYEIPGYTSKNFDMLEEVQSPESALSSKSKSEIF 360
Db 301 INKPSITNTSWEMIKLEAIMKYEIPGYTSKNFDMLEEVQSPESALSSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
Db 361 LPDDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMFKITNYLKVGFPADVRSTINSGPGVYTGAYQD 480
Db 421 GTDFNTMTKIFSDKLASISNEDNMFMFKITNYLKVGFPADVRSTINSGPGVYTGAYQD 480
QY 481 LLMFKDNSTNIHLEPELNFPPKTKISQTEQITSLWSNQARAKSQFEYKKGYPF 540
Db 481 LLMFKDNSTNIHLEPELNFPPKTKISQTEQITSLWSNQARAKSQFEYKKGYPF 540
QY 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKEYIHYIYVLOQDKISYEASCNLF 600
Db 541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKEYIHYIYVLOQDKISYEASCNLF 600
QY 601 DPYSSILYQKNEGSETAYYYVADAEIKEIDKIRIPYQISNKRNIKLTFFIGHGSEFNT 660
Db 601 DPYSSILYQKNEGSETAYYYVADAEIKEIDKIRIPYQISNKRNIKLTFFIGHGSEFNT 660
QY 661 DTFANLDVDSLSSEITILNLAADISPKYIEINLLGCMNFSYSIYAEETYPGKLLKIK 720
Db 661 DTFANLDVDSLSSEITILNLAADISPKYIEINLLGCMNFSYSIYAEETYPGKLLKIK 720
QY 721 DRVSELMPSISODSITVSANQYEVRIENERGKEILDHSGKWNKESIIKDSSKEYISF 780
Db 721 DRVSELMPSISODSITVSANQYEVRIENERGKEILDHSGKWNKESIIKDSSKEYISF 780
QY 781 NPENKIIIVKSKYLHELSTLLOEIRNANSDDIDLEKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPENKIIIVKSKYLHELSTLLOEIRNANSDDIDLEKVMLTECEINVASNIDRQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEPKLIESDSISLYDLKHQGLDSDSHFISFEDISKTENGFRIRF 900
Db 841 IBEAKNLTSDSINYIKNEPKLIESDSISLYDLKHQGLDSDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSIFETEKEIPEYATHISKEISNKTIDTFDNNVGLVKVKNLDAAEVNTLN 960
Db 901 INKETGNSIFETEKEIPEYATHISKEISNKTIDTFDNNVGLVKVKNLDAAEVNTLN 960
QY 961 SAFFIQSLIEYNTKESNLNSVAMKVQVYVQYVQALFSTGLNTIITDASKVVELVSTALDETID 1020
Db 961 SAFFIQSLIEYNTKESNLNSVAMKVQVYVQYVQALFSTGLNTIITDASKVVELVSTALDETID 1020
QY 1021 LPLTLEGLPIIATIIDGVSGLGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIYV 1080
Db 1021 LPLTLEGLPIIATIIDGVSGLGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIYV 1080
```

```
QY 1081 SALGIASGSILLVPLAGISAGIPSLVNNELIQQKATKVIDYFKHISLAETEGAFITLD 1140
Db 1081 SALGIASGSILLVPLAGISAGIPSLVNNELIQQKATKVIDYFKHISLAETEGAFITLD 1140
QY 1141 DKIIIMPQDVLSEIDFNNSITLKGCEIMRAEGSGHTLTDIDHFPSSPSITYRKQWL 1200
Db 1141 DKIIIMPQDVLSEIDFNNSITLKGCEIMRAEGSGHTLTDIDHFPSSPSITYRKQWL 1200
QY 1201 SIYDVANIKKIDFSDKDLAVLPNAPRVFGYEWGTPGFRSLNDGDKLDRIDHYEG 1260
Db 1201 SIYDVANIKKIDFSDKDLAVLPNAPRVFGYEWGTPGFRSLNDGDKLDRIDHYEG 1260
QY 1261 QFYWRYFAFIADALITKLKPRYEDTNVRLNDGNTRSPFIVPITTEQIRKNLSYFYGSG 1320
Db 1261 QFYWRYFAFIADALITKLKPRYEDTNVRLNDGNTRSPFIVPITTEQIRKNLSYFYGSG 1320
QY 1321 GSYSLSPYNNIDNLVENDTWIDVDNVVKNITTESDEIQGELIENILSKLNIEDN 1380
Db 1321 GSYSLSPYNNIDNLVENDTWIDVDNVVKNITTESDEIQGELIENILSKLNIEDN 1380
QY 1381 KIILNNHTINPYGDNESNRFISLTFSLEIDNIIIEIDLVSKSYKILLSGNCKLIENS 1440
Db 1381 KIILNNHTINPYGDNESNRFISLTFSLEIDNIIIEIDLVSKSYKILLSGNCKLIENS 1440
QY 1441 SDIOQKIDHIGFNGEHQKVIYFVYIDNETKNGFTIDYSKKEGLFTAEPNSNIIIRNIYMP 1500
Db 1441 SDIOQKIDHIGFNGEHQKVIYFVYIDNETKNGFTIDYSKKEGLFTAEPNSNIIIRNIYMP 1500
QY 1501 DSNLLFYSSKDLKDIRIINKGDVKLLIGNYFKODMKVLSFTTIEDTNTIKLNGVYLDEN 1560
Db 1501 DSNLLFYSSKDLKDIRIINKGDVKLLIGNYFKODMKVLSFTTIEDTNTIKLNGVYLDEN 1560
QY 1561 GVAQILKFMNNAKSALNTSNLMNPLESINIKNIFNNLDNPIERFILTNTIISGNSIG 1620
Db 1561 GVAQILKFMNNAKSALNTSNLMNPLESINIKNIFNNLDNPIERFILTNTIISGNSIG 1620
QY 1621 QFELICDKDKNIQPIYFNFKIKETSYTLVGNRQNLIVPEPSVHLDDSGNISSTVINFSQK 1680
Db 1621 QFELICDKDKNIQPIYFNFKIKETSYTLVGNRQNLIVPEPSVHLDDSGNISSTVINFSQK 1680
QY 1681 YLYGIDRYNVKVIITAPNLYT 1700
Db 1681 YLYGIDRYNVKVIITAPNLYT 1700
RESULT 2
Q9EXR0
ID Q9EXR0 PRELIMINARY; PRT: 2367 AA.
AC Q9EXR0;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Toxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RC STRAIN=8864;
RP SEQUENCE FROM N.A.
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19891.1; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding_1; 17.
```











Db 763 DYFGKWSNTDLIAEQISKNKYVYVWNEVENTLSARVEQLNKVAEPAKDI-----NSIIQT 817  
Qy 816 EKKVWLTECEINVASNIDRQIVEGRIERAKNLTSDSI---NYIKNEFKLIESDSLYD 871  
Db 818 TNNQELKOSLVNTYADLTITLYSELKEDDIPFELDNIOIKERILNEISRHDHSDNIID 877  
Qy 872 LKHQGLDSDHIFSPEDISKTEGFRIRPINKETGNSIFIEKEIFSEYATHISKEISN 931  
Db 878 FYQKNISNNMIILPDSIIKEKDYVNVKLANKITGETSVIKTYSDSLWNFTNKYKIVDD 937  
Qy 932 IKDTIFDNNVGLVKVKNLDAAEVENTLNSAPFIOSLIEYNTTKESLNSLVAMKVQVYA 991  
Db 938 IKGIIVVDINGEFIKKAEFEQNPSSLNSAMQLLDIDYKPYTEBILTNMNTSLKVQVYA 997  
Qy 992 QLFSTGLMTITDASKVVELSTALDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSET 1051  
Db 998 QIFQLSICAIQATEIIVIIISDALNANFNILSKLVGSSVASVIIIDGINLIALTELKVV 1057  
Qy 1052 NDPLLRQIEAKIGIMAVNLTAATAIVTSALG---IASGFSILVPLAGISAGIPSLVN 1108  
Db 1058 KTNPERKLEAKGVMSIGFLESLSLGLGATAVSEILGVISVPVAGILVGLPSLVN 1117  
Qy 1109 NELIQQDKATKVIDYFKHISLAETEGATLDDKLIIMPQDDLVLSIEDPNNSITLGKCE 1168  
Db 1118 NILVGEKYNQILDYFSKPYFIVGNKPFIS-IQDNIIIPYDDIAITELNFKYKGYAK 1176  
Qy 1169 IWRABEGSGHTLTDIDHFFSPSPITYRKPMLSIYDVLNIIKKEIDFSDKMLWLPNAPNR 1228  
Db 1177 ISGLKVLGTHIGENIDHYFAPSUDHYE-LSIYPALKLNDTNLP-KGNVVLLPSGLNK 1234  
Qy 1229 VFGYEMGWTPGPRSLDNDGTLLDRIDHY-----EGQFYWRYFAFIADALITKLKPYE 1283  
Db 1235 VYKPEISALAGANSQEGGVEVLNIRNYVDSNGNTKPPMKYEPF-EYSFYSRVEYF 1293  
Qy 1284 DTNRVINDGNTRSFIVPITTEQIRKNLSYSFYGGSGYSLSLSPYNNIDNLNVENDT 1343  
Db 1294 DTKVNVILDENKTLIIPLVITIDERNKISYEILGDGGYVILPVNQTINIVSNKNDI 1353  
Qy 1344 WVIDVNVKNITTESDEIQGELLENILSKLNIEDNKIILNNHTINFYGDINESNRFIS 1403  
Db 1354 WNFDSYIVKSKIEEDNKPVLDGFINNIPSTLKVSDNGFKIGKQFIS----IKNTPRAIN 1409  
Qy 1404 LTFSELEDINIIIEIDL-VSKSYKILLSGNCMKLIENSSD-IQQKIDHIGFNGEHQYIF 1461  
Db 1410 LSFKINNI-VIVSYLNAHKSNTIISDLDNDIKNNFDNLDNINYLGLSISDNTIN 1468  
Qy 1462 YSIDNETKNGFIDYSKKEGLTFAEPNESIIRNYPDSNNLPYSSKDKLDIRINK 1521  
Db 1469 CIVRNDVYMEGKI-----FLNEK--KLVPFQNELELHLYDS-----VNK 1506  
Qy 1522 GDVKLLIGN-----YFKDDMKVSLSPFTIEDT-----NTIKLVGVLDENGVAQI 1565  
Db 1507 -DSQYLINPINNVVKYKDYIVETGFTLINSTENKYSLYIENKNMLKGLYL-ESSVFPT 1564  
Qy 1566 LKFMNNAKSALNT-SNSLNFPLESNIK---NIPYNNLDPN---IEFILDN--PIISGS 1616  
Db 1565 IQDKIYSKEVNDYLSLKKEFTVNIQCPMIVSGVDENNRYLEYMLSTNNKWIING- 1623  
Qy 1617 NSIQGFELICDKKNIQPYFNFKIETSYTYLVGNRONLIVEPSYHLDGNSISSTVIN 1676  
Db 1624 ---GYWE---ND-----FNNYKIVDF-----EKCNVIVSGSKLNSGDLADT-ID 1662  
Qy 1677 PSQKYLIGIDRVNKKVIIAPNLYT 1700  
Db 1663 VLDKDLLENL--YIDSVIIIPKYYT 1684

RESULT 7

O68653

ID O68653

AC O68653;

DT 01-AUG-1998 (TREMELrel. 07, Created)

DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)

PRELIMINARY;

PRT;

698 AA.

01-DEC-2001 (TREMELrel. 19, Last annotation update)

Truncated toxin A.

OS Clostridium difficile.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI\_TaxID=1496;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=CCUG 20309;

RX MEDLINE=99304497; PubMed=10376209;

RA Song K.P., Bai X.L., Chang S.Y.;

RT "Nucleotide and peptide sequences of the open reading frame encoding a

truncated toxin A gene of Clostridium difficile strain CCUG 20309."

RL DNA Seq. 10:93-96(1999).

DR EMBL; AF053400; AAC08437.1; -- 021EB268A3BDECSE CRC64;

SQ SEQUENCE 698 AA; 80682 MW; 021EB268A3BDECSE CRC64;

Query Match 22.0%; Score 1911; DB 2; Length 698;

Best Local Similarity 52.9%; Pred. No. 7.6e-67;

Matches 371; Conservative 136; Mismatches 186; Indels 8; Gaps 5;

Qy 1 MNLVYKQLOKQWVYVKERIOEDYVAIILNALSEYHNHNSSESVVEKYLKLDINNLTNYL 60

Db 1 MSLISKBELIKLAY-SIRPRENYKYTVLTNLDYENKLTNNENKYLQKKLNSIDVFM 59

Qy 61 NTYKSGRNKALKKFKPEVLTWVLELKNLSLTPVEKNLHFWIGQINDTAINYNOWKD 120

Db 60 NYKNSRNRLSNLKKDLKEVILIKNSNTPVEKNLHFWIGVEVSDIVLEIKQWAD 119

Qy 121 VNSDYTVKVVYDSNAFLINTLTKTIVESATNNTLLESFRENLDPEFDYKFKRMEIY 180

Db 120 INAEYNIKLYDSEAFVLTWTKAIVESSTTEALQLLEEBEQNFQDNMKFYKRMFEY 179

Qy 181 DKQHFIDYKSOIEEPEFIIDNIIKLYLSNEYSKDLALNKYIEESLNKITANGNDI 240

Db 180 DRQRFNYKYSQINKPFTVLTIDDIISYLVSEYNRDETLESYRTNSLRKINSNHGIDV 239

Qy 241 RNLEKFADEDLVRVLYNOELVERWNLAAASDLRLISMLKEDGGVYLDVLDLPGIOPDLFS 300

Db 240 RANSLFTEQELLINYNQELNRLNGLNLAASDIVRLALKNFGVYLDVDMFGIHSDLFT 299

Qy 301 INKPDSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVQSFALSCKSDKSFIF 360

Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYNNTYSENFDFKLDQQLKDFKLIIESKSEKSFIF 359

Qy 361 LPDDIKVSPLEVIAFANNNSVINQALISLKDSCSDLVINOIKRKYKILNDNLNPSINE 420

Db 360 SKLENLVSDLEIKIAFALGVSINQALISKSGSYLTNLVIEQIKRKYQFLNQHLPATES 419

Qy 421 GTDFNTWKIPSDKLASTSNEDNMFMKINTYLVKGFAPDVRSTINLSGPGVYTGAYOD 480

Db 420 DNNFTDTTKIFHDSLPNSATENSMTLKAPYLOVGFMPARSTISLSGFGAYASAYD 479

Qy 481 LLMPKONSTNHLLEPELNFPEPKTKISQTEQIETSLWSFNQARAKSQFEYKKGTYFE 540

Db 480 FINQENTIEKTLKASDLIEFKFPENNLSQLTEQINSLSFSDQASAKYQFEKYKDYTG 539

Qy 541 CALGEDDNLDAQNTVLDKDY-VSKLILSS--MKTRKYEYHYIVOLQGDKISYEASCNL 597

Db 540 GSLSEDEVDNFNKTALDKNYLNNKIPSNVVEBAGSKNYVHYIQLQGDIDISEATCNL 599

Qy 598 FSKDPYSILYQKNIEGSETAYVYVAD--AEIKEDKYRIPYOISNKRNTKLPFIHGK 655

Db 600 FSKPKKSIILQRRN--NESAKSYFLSDGSEIIELNKRIYPERLNKPKVTFIHGK 657

Qy 656 SEFNTDTFANLDVDSLSEIETILNLAKADISPKYIEINLL 696

Db 658 DEFNTSEFARLSDVSLSEISFLDTIKLIDISPKNVEVNL 698

RESULT 8

Q9EXQ8

ID Q9EXQ8

PRELIMINARY;

PRT;

698 AA.

```
AC Q9EXQ8;
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-WAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Truncated toxin A.
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RV [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RT "The pathogenicity locus of Clostridium difficile, strain 8864.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A3011301; CAC19893.1; -.
SQ SEQUENCE 698 AA; 80799 MW; 66D9B21798D314BB CRC64;

Query Match 21.8%; Score 1893; DB 2; Length 698;
Best Local Similarity 52.4%; Pred. No. 3 8e-66;
Matches 367; Conservative 138; Mismatches 188; Indels 8; Gaps 5;

QY 1 MNLVNAQLOKMYVYKFRIOEDYVAIALNALBEYHNMSSESVVEKYKLKDKDINNLTNYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKVTNLNDEYNKLTNNNNENKYLQKLKLNESIDVFM 59

QY 61 NTKYSGRNKALKKPKYELTMEVLEKNNSLTPVEKNLHFIWIGGQINDTAINYNQKDV 120
DB 60 NKYKSSRNALSNLKKDLKEVILKNSNTSPVEKNLHFIWIGGQINDTAINYNQKDV 119

QY 121 VNSDYTVKPYDGNAPFLINTLTKKTIVESATNTNLESFRENLDNDFYKFKRMEIYY 180
DB 120 INAEYNIRLWYDEAFVNLTKKAI VESSTTEALQLEBEIQNPQDNMFKYKRMFIY 179

QY 181 DRQKHFIDYKSOIBENPEFIIDNIITKYLSEYKDEALNKYIIBESLNKITTANNNDI 240
DB 180 DRQKHFIDYKSOIBENPEFIIDNIITKYLSEYKDEALNKYIIBESLNKITTANNNDI 239

QY 241 RNLEKFADEDLVRLNQEELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQPOLFS 300
DB 240 RANSLFTEOELLNIYQNELNLRGNLAAASDIIRLLALKNFGVYLDVDMVLPGLHSDLFKT 299

QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSIF 360
DB 300 ISRPSIGLDRWEMIKLEAIMKYKEYINNTYSENFDFKLDQQLKDNFKLIESKSKCIF 359

QY 361 LPLDDIKVPLEKIAFANNVINQALISLKDYSYCSLVINQIKNRYKILNDNLNPSINE 420
DB 360 SKLENLNSDLKIRFALGVSINQALISKGSYLTNLVIEQIKNRYQFLNQHLPALIES 419

QY 421 GTDFNTMTKIFDKLASINEDNMFMKIITNLYKVGAPDVRSTINLSGPGVYTGAYQD 480
DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPLYQVGFMPPEARSTISLSGPGAYASAYD 479

QY 481 LLMFKDNTNIIHLEPELRNFPFKTKISQLETSLSWSPNQARAKSQPEYKGYFEGA 540
DB 480 FINLQENTTEKTLKASDLIEFKFPENNLSQLETSLSWSPNQARAKSQPEYKGYFEGA 539

QY 541 GALGEDDNLDFQAONTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKIISYASCNL 597
DB 540 GSLSSEDEYDFNKNTALDKNLLNNKIPSNVVEAGSKNYHYIILQLODDISYEATCNL 599

QY 598 FSKDPVSSLYQKNTLEGSTAYYYVAD--ARIKEIDKYRIPYQISNENKILTEIGHCK 655
DB 600 FSKNPKNSIIIQRM--NESAKSHFLSDDGESILENLKRYIPERLNKNEKVKVTFIGHCK 657

QY 656 SEFNFTDTFANLDVDSLSSEIETILNLAKADISPKYIEINLL 696
```

```
DB 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLDISPKNVEVNL 698

RESULT 9
ID O86141 PRELIMINARY; PRT; 697 AA.
AC O86141,
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TCdA protein (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Soehn F., Leukel P., Weidmann M., Eichel-Streiber C.V., Braun V.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RV [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y10689; CAA71690.1; -.
FT NON TER 1
SQ SEQUENCE 697 AA; 80680 MW; CD650D3372D65B73 CRC64;

Query Match 21.7%; Score 1887; DB 2; Length 697;
Best Local Similarity 52.4%; Pred. No. 6.5e-66;
Matches 366; Conservative 137; Mismatches 188; Indels 8; Gaps 5;

QY 3 LVNKAQLOKMYVYKFRIOEDYVAIALNALBEYHNMSSESVVEKYKLKDKDINNLTNYL 62
DB 2 LISKEELIKLAY-SIRPRENEYKVTNLNDEYNKLTNNNNENKYLQKLKLNESIDVFM 60

QY 63 YKSGRNKALKKPKYELTMEVLEKNNSLTPVEKNLHFIWIGGQINDTAINYNQKDV 122
DB 61 YKSSRNALSNLKKDLKEVILKNSNTSPVEKNLHFIWIGGQINDTAINYNQKDV 120

QY 123 SDYTVKPYDGNAPFLINTLTKKTIVESATNTNLESFRENLDNDFYKFKRMEIYYDK 182
DB 121 AEYNIRLWYDEAFVNLTKKAI VESSTTEALQLEBEIQNPQDNMFKYKRMFIYDR 180

QY 183 QKHFIDYKSOIBENPEFIIDNIITKYLSEYKDEALNKYIIBESLNKITTANNNDIRN 242
DB 181 QKHFIDYKSOIBENPEFIIDNIITKYLSEYKDEALNKYIIBESLNKITTANNNDIRN 240

QY 243 LEKFADEDLVRLNQEELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQPOLFSIN 302
DB 241 NSLFTQOELLNIYQNELNLRGNLAAASDIIRLLALKNFGVYLDVDMVLPGLHSDLFKTIS 300

QY 303 KPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSIFELP 362
DB 301 RPSISGLDRWEMIKLEAIMKYKEYINNTYSENFDFKLDQQLKDNFKLIESKSKCIFSK 360

QY 363 LDDIKVPLEKIAFANNVINQALISLKDYSYCSLVINQIKNRYKILNDNLNPSINEGT 422
DB 361 LENLNSDLKIRFALGVSINQALISKGSYLTNLVIEQIKNRYQFLNQHLPALIESDN 420

QY 423 DFNMTKIFDKLASINEDNMFMKIITNLYKVGAPDVRSTINLSGPGVYTGAYQDL 482
DB 421 NFDFTDTTKIFHDSLFNSATAENSMFLTKIAPLYQVGFMPPEARSTISLSGPGAYASAYDFI 480

QY 483 MFKDNTNIIHLEPELRNFPFKTKISQLETSLSWSPNQARAKSQPEYKGYFEGA 542
DB 481 NLQENTTEKTLKASDLIEFKFPENNLSQLETSLSWSPNQARAKSQPEYKGYFEGA 540

QY 543 LGEDDNLDFQAONTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKIISYASCNLFS 599
DB 541 LGEDDNLDFQAONTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKIISYASCNLFS 599
```

```
Db 541 LSENEVDNFKNVATLDKNNLNNKPNVNNVERAGSKNYYHYIIQGGDISYEATCNLFS 600
Qy 600 KDPYSSILYQKNIQEGSETAYYYVAD--AEIKEDIKYRIPYQISNKRNIKLTFIGHKSE 657
Db 601 KNPKNISIIQRM--NESAKSHFLSDGESILELNKYRIPERLKNKKEKVKYTFIGHGKDE 658
Qy 658 FNTDTFANLDVDSLSSEIETILNLAKADISPKYIEINLL 696
Db 659 FNTSEFARLSVDLSLSNEISSFLDTIKLIDISPKNVEVNL 697

RESULT 10
Q9FCX5 PRELIMINARY; PRT; 553 AA.
AC Q9FCX5;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Toxin A (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C34;
RX MEDLINE=20402122; PubMed=10931294;
RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
RA von Eichel-Streiber C.;
RT "A chimeric ribozyme in Clostridium difficile combines features of
RT group I introns and insertion elements.";
RL Mol. Microbiol. 36:1447-1459(2000).
DR EMBL; AJ131844; CAC03681.1; -.
FT NON_TER 553
SQ SEQUENCE 553 AA; 63913 MW; DDB4551A6D8C3B25 CRC64;

Query Match 16.9%; Score 1469; DB 2; Length 553;
Best Local Similarity 51.7%; Pred. No. 9.1e-50;
Matches 287; Conservative 106; Mismatches 160; Indels 2; Gaps 2;

Qy 1 MNLVNAQLOKRVYKFIQDEYVAIINLALNEEYNNMSESSVVEKYKLKIDNNITDNYL 60
Db 1 MSSIISKELIKLAY-SVRPRENEYKLTILNLDYNNKLTNNENKYLQKLKLNESIDVFM 59
Qy 61 NTYKSGNKALKKEVILTMEVLELKNLSLTPVEKNLHFVIGQINDTAINYNQWID 120
Db 60 NKYKNSRRNALSNLKDKILKEVILLIKNSNTPVEKNLHFVIGVSDIALEYIKOWAD 119
Qy 121 VNSDYTVFVDSNAFLINTLKKTIVESATNNTLESFRENLDPEFDYNNKFKRMEIY 180
Db 120 INAEYVNLWDSEAFVNTLKAIVESTTEALQLLEEIQNFQFDM-KFYKKRMEFIY 178
Qy 181 DKQHFIDYKSIQIEENPEFIIDNIITKYLNEYSKOLEALNKYIEESLNKITANNNDI 240
Db 179 DRQKRFINYKIQNKPTPTIDDIKSHLVSEYNRDETLLSYRTNSLRKINSNHGIDI 238
Qy 241 RNLEKFAEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDLPKS 300
Db 239 RANSLFTQELLNYSQBLNRGNLAAASDIVRLALKNFGGVYLDVDMVLCIHSDLFKT 298
Qy 301 INKPDISTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMDEEVORSFESALSKSKSEIF 360
Db 299 IPRPSIGLDRWEMIKLEAIMKYKYYINNYSENFDKLDQQLKONFKLIIESKSEKIF 358
Qy 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINGIKRKYKILNDNLNPSINE 420
Db 359 SKLENLNSDLEIKIAPALGVSINQALISKQGSYTNLVIEQVKNRYQFLNQHLPATES 418
Qy 421 GTDFTNTWKIPSDKLASISNEDNMFMFKITNYLVKGFAPVRSTINISGPGVYTCAYQD 480
Db 419 DNNFTDTTKIFHSDLSFNATAENSMLFKIAPYLVQGFMPPEARSTISLSGFCAYASYD 478
```

```
Qy 481 LLMFKDNSTNIHLPELNRNFEPPKTKISQITQSEITSLMSFNQARAKSQFEYKKGYPF 540
Db 479 FINLQNTIETKLKASDLIEPKFPENNLSQLTQEINSLWSFDQASAKYQIERYVDYTG 538
Qy 541 GAGEDDNLDPAQNT 555
Db 539 GSPSGDNGVDNFNKT 553

RESULT 11
Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
AC Q9ZGR4;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Putative cytotoxin.
GN I7095.
OS Escherichia coli O157:H7.
OC Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
RT of Escherichia coli O157:H7.";
RL Nucleic Acids Res. 26:4196-4204(1998).
DR EMBL; AF074613; AAC70163.1; -.
DR InterPro; IPR006473; YOP_T.
DR TIGRFAMs; TIGR01586; YopT_cys_prot; 1.
SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BDSFD1 CRC64;

Query Match 5.8%; Score 501.5; DB 2; Length 3169;
Best Local Similarity 20.1%; Pred. No. 3.8e-11;
Matches 421; Conservative 331; Mismatches 723; Indels 619; Gaps 104;

Qy 14 YVKFRI---QDEYVAIINL--BEYNNMSESSVVEKYKLKIDNNITDNYLTKSGR 68
Db 160 YIKIRKTRGAEQDTTITQSLIINELLGVDRNTI-PPQKISELNDIHSYENMQIKNSR 218
Qy 69 N--KALKKFKYELTMEVLELKNLS-----LTPVEKNLHFVIGQ 106
Db 219 KGIBILVKQGBLLSLINDNKGKQLSDNASKIINLLGIEYQSHKVDIEPPIHAWVAGA 278
Qy 107 INDTAINYINQWKNVSDYTVKVVDSNAF-----LINTLKKT----- 144
Db 279 PPDNTFSYITAFNTYKDYTYLLWDIDNPAFGAAGKPSGLKNIAINYAIRLRNTNPHLAE 338
Qy 145 -----IVESATNNTLE--SFRENLDPEPDY-----NKFYKMEIYIDKQKHFI 187
Db 339 EMNEVILKIQNIQNETIEPKETRELRKLENNYKSLTSETKEKNVFNLESWIGNQDNYF 398
Qy 188 DYKSIQIEENPEFI-----IDNIK--TYLSNEYSKOLEALNKYIEESLNKITANN-- 238
Db 399 TYCISNGISNTDDISRLDFTLVNLKLSPEVQNDPKFTVEKNKRDIDLLKNTISQKGDGF 458
Qy 239 ---DIRNLEKFAEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQP 295
Db 459 QLRDINTLESPPKPDQVFFYQOEWLLRNWYAAASQVRINILKSYGGIYTTDILPAYSD 518
Qy 296 DLPKSINKPDSITNTSWEMIKLEAIMKYK--EYIPG--YTSKNFDMDEEVORSFESALS 351
Db 519 KVSQIINE-KSDDKRFPEDLKLRIISLSEIUSLKGKYSIKH-DGLDETTLNQLNNIL- 575
Qy 352 SKSKSKSEIFLPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINGIKRKYK 395
Db 576 SEIEK----LTIDY-FKPVETKVVDRDTFKFKRYQKWTENTWIRGNMNFMLTHGSKC 630
```



Db 856 LSTYNSTELNKLNFSSKDDDELNVESKYQDENIFIEKNKIPFDDIIOIKDIELYNKKTNAI 915  
Qy 32 BEYHN-----NSESSEVVEKYKLKD--INNLE-----TDNLYLTYK-----KSGRNKAL 72  
Db 916 KULNNAINGSMMNLSDISVMKNKGDIINRUSQSYLYIQTNFIDYIEKIFUKONLNG 975  
Qy 73 KKFKEYLTWEVL-----ELKQNSLTPVEKNL-----HPIW--IGQIND 109  
Db 976 BEIENRLSNTYNNVELKIBAEQNEKYKGLKENINTYDDTFLEKLGIDNVEWELKIELNG 1035  
Qy 110 TAINY-----INQWQVNSDYTVKFFVDSNAFLNLTAKTIVESATNNTLESFR 158  
Db 1036 LNVNTNLOANIDTLLIIRPYID-HIDHIIISLESLEKHNIEKNIKKIVIPNLEBLKDFIQTK 1094  
Qy 159 ENLNDPEFDYKFKRMEIIVDKQHFIDY-YKSOIRENPEFIDNIIKTLYLSNEYSKD 217  
Db 1095 FNTNDIKLQHNLIIRI-----DNR-----DYNHMKLEEKEDLPKNI-----NDKKEE 1140  
Qy 218 LEALNKYIEESLNK-----ITANNNGDI-----RNLEKFADE 249  
Db 1141 IEKXLKLEENNTKEMENNTISYGIKKKNLIDYESMSSLLAKSIITTDENLYKLQNV 1200  
Qy 250 DLVR-LYNQELVERWNLAASDILRISMLKEDGGVLOVDILP-----GIQDLFKSINKPD 305  
Db 1201 DEFRLLYELILTEERINEKIQKEASVERVEIELYKEKILSMNNSIKEDISK-----1254  
Qy 306 SITNTSWEMIKLEATMKYKEYIPGYTSKNFMDLDEEVQSPESALSSKSKSEIPLPDD 365  
Db 1255 -LTPKYNEYDECITKENNIKELYT-XSSSLEE-----SCRDKN-----MDI 1296  
Qy 366 IKV--SPLE--VKIAPANNVINOALISLKDSCYSDLVINOJK-----NRYKILN 411  
Db 1297 IKTNSVLDEYLTQSIQNNNDIKNSLASLKNWYA---LQSIKLDVATKYILDNSYKC-- 1351  
Qy 412 DNLNPSINEGDTFNTMTIFSDKLASISNEDNMFMKITYNLYKVGFPADVRSTINLSGP 471  
Db 1352 -----EDHARDLEMELEKESALSKN-----IK-----LKTEKAEVEYRNKV-----1386  
Qy 472 GYVTGAYQDMLLFKDNSTNIHLLEPELNFPEPKTKISOLTEQETSLWSP--NQARAKS 529  
Db 1387 -----LGSEKHESIDY-----IRNIEKIQDISVIESDIMTICIQNAYDNKTKSML 1432  
Qy 530 OPEEYKKG-YFEGALGEBDNLDPFQNTVDKDYVSKILSSMKTRNKEYIHYIVLOQDK 588  
Db 1433 HFQNVHRGDLILGNKN-----QGVIKPAESGNI-----SEY-----EQON 1472  
Qy 589 IS-YEASCNLFK-----DPYSSILYQKNIEGSETAYYYYVADAEIKEDK-----633  
Db 1473 INDYQOKCKTYSEASDNYEISYRKD-----SLLEFEKKITNILDVLI 1517  
Qy 634 YRIPQISNKN-----IKLT-FIGHKSEFNTDTFANLDVSLDSEIETILNLA 684  
Db 1518 FNMKKTLENKQSVNNILEGMKLTASIIINEKSHVLSEKIRVLRNQKNIKODBELLNNEKS 1577  
Qy 685 DISPKVIEINLGCNMFYSIYAEETYPCKLLKIKD-RVSE-----LMPSI 730  
Db 1578 KIAYQOFEL-----YMGLEYAIPDINRLHDKAERMFPSKAQFLMKPWL 1620  
Qy 731 SODSITVSANOYEVINEBEGREILDHSGKWNKEESIIKOISKEYISFNPKE-----N 785  
Db 1621 DMSLILEKNKLEELK-----NKEHDYMEI---EYIQDEKHIKYELT 1660  
Qy 786 KIIVSKYLHELSTLQELIRNANSSDIDLEKXVMLTECEINVASNIDROIVEGRIBAK 845  
Db 1661 KLIEVIQFVDNIENTLEKRYKYEQGNLE-----NVYNSNN--IKDRIEETK 1706  
Qy 846 NLTSPISINVKNEFKLIESISDLYDLKHQGLDSDHFSFEDISKTENGFRIRINKET 905  
Db 1707 NLSVLVN-----IFSIKNSIYLLKHS-----VKMEDLNS-----1737  
Qy 906 GNSIFETEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKKVNLDAAEHVNTLSAPFI 965

Db 1738 ----YIKQNDIYDEF-----MESYNLLQKTIIESSN-----DDIEYEELKQVNRNIEKPEI 1785  
Qy 966 QSLIEYNTTKESNLNSVAMKVQVYAOULFSGLNTITDASKVVELVSTALDETIDLLPTL 1025  
Db 1786 QLV-----KAEDMKLYNSIKNDVYNKLVYIKWMLLELDNRCKSVDTILNEGDFPNKC 1841  
Qy 1026 SEGPIIATIIDGV--SLGAAI-----KELSETNDPLLRQIEBAKIGIAMVNIATASTAIVT 1080  
Db 1842 AKESNVSGD-DDNVVYNELNKAINKYKEIHEKSNFVKNEAESLFGIIVKSSNIIGMKIIT 1900  
Qy 1081 SALGIASGFSILLVPLAGISAGISPLVNNEL-----ILOQKATKVIDYPKHI 1127  
Db 1901 -GLG-----LELKEDVDLGTWLSLSSLHFHTASINKLYSTIESDVKVNDCLKS- 1950  
Qy 1128 SLAETEGAFLLDDDKIIMPQDDLVSEIDFNNSITLCKEIRWAEAGSGHTLTDIDH 1187  
Db 1951 SLDIVKYSFDEKKK--RQINSIMEELNVAHITI-KIE-----LANMINDS 1995  
Qy 1188 FSSPSITYRKPLWSIYDVLNIIKKEKIDPSKOLMVLPNAP-NRVFGYEMGW-----1236  
Db 1996 RNKISVVLNKIYVSINKIKVKEMTCDSDSYHMIWKEGEYDKLKEYYKNYNEEKLATNE 2055  
Qy 1237 --TGFPSLNDGDKLLDRIRDHYEGQPYWTFYAFIADALITKLKPRYEDTVNRINLGN 1294  
Db 2056 LNTNKLKENSCKLLDLQNTMKNTIEE-----SPKKVIEDIKKSYDEINGRI---GN 2108  
Qy 1295 TR--SPFIVPVTITQI-----RKNLSYSFYS-- 1319  
Db 2109 TEMDAEAINVVEBELVKRQNKASWYTSLLGKNTKMSFKKLVDVRQKISQSFDSMK 2168  
Qy 1320 --GGSYSLSLSPYNNIDNLVENDTWVIDVNVKNITIESDEIQG-----ELIENILS 1373  
Db 2169 LNFNMINKWINKINHFDKQINN-----YSLGNVEKMMYINDSNEKTSVIELIDNLTK 2224  
Qy 1374 KLN-----EDNKIILNNTINPYGDNES-----NRFISLTF 1407  
Db 2225 KISINNSNNNNNNNNNNNNNDIINFINTKENLITNYENLGKCKXDLNIIYLKIFPS 2284  
Qy 1408 ILEDINIIE-IDLVSKSYKILLSGNCMKLEN-----SSDIQOK-----1 1447  
Db 2285 KLEEIEKSEKIFYSKAFNNVVKYQOGKSDNINRMNIIIEEKETILNNKNTL 2344  
Qy 1448 DHI-GFNGEHQKIFYS---YIDNETKY-----NGFIDYKCKGLFTAESNESIIRNI 1497  
Db 2345 DTIKEFNKAYDNILFHMREDYLBESKYDVGRKLNIIYIEY-----IYQIRKNSNNLINNF 2399  
Qy 1498 Y-MPDSNNLFYSSKDLK-----IRINKG-----DVKL-----LIGNYF---KDDMK 1537  
Db 2400 YDIRNTGYIMIHETEKLNIDINEYMNKINKNINNTNKEFDITLRNKTNNYFVNNHBEIK 2459  
Qy 1538 VLSFTIEDTNTIKLN-GVILDENGVAQILK-FMNNKASALNTSLSMNFLESINIK-NI 1594  
Db 2460 KVIDNFWNHLDTIKANFTFLPQNEKVLIIKSYLNDIK---NIMNELMR-KEQVDVSIQT 2515  
Qy 1595 FYNLDPNIEPILDTNFTIISGNSISGQFELICD-----KQKNIOPYFINFKIKETS 1645  
Db 2516 CTRNLDIELEKI-----NKIKQNNALTDQVENVFLSKDKYKNEFLLTKDILKRV 2565  
Qy 1646 YTLVGNRQNLIVBP-SYHLDDSGNISSTVINFSQKLYGIDRYVNVK 1692  
Db 2566 QDKY--QEMNKIYRTLTVTNTNNEN-----KFALKYLSNVNAPFINEI 2605  
RESULT 13  
ID O82916 PRELIMINARY; PRT; 3169 AA.  
AC O82916;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Toxin B.  
GN TOXB.  
OS Escherichia coli O157:H7.











```

Qy 707 -----ABETPGKLLKIKIDRVSELMPSISQDSITVSANOYVRINEGKR 752
Db 1303 SDIRKNSLKIIQDFSEESVINDIKELEKNVLE-----SQNNNT-DINOYLSKI-ENIY 1354
Qy 753 BILDHSGWINKEESIIKDISSKEYISFNPKENKIIIVSKYLHELSTLLOERNNANSSD 812
Db 1355 NILK-----LNKIKIIDKV--KEYTDIERKNK-----KINAEULS-----NSBKII 1394
Qy 813 IDLEKVMLTECEINVASNIDRQIVEGRIEAKNLTSDSIN-----YIK-----856
Db 1395 TOLKENSSLEKQSKISTIDNNYVSECINITNLKTYIVNEKNWINTYFKNABEYNQNV 1454
Qy 857 -----NEFKLIESISDLYDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NSI 909
Db 1455 SLNFNNIEMADTKSQYIILNKKNGTNTDYNIKELKHKKSNNYK-----DEAGKNTQ 1509
Qy 910 FIETEKEIFSEYATHI-----SKEISNIKOTIFDNNVNGKLVKKNLDAAEVNT---958
Db 1510 EIKKNKELFEYEQEVTVLLNKYYAVELKNKFDKT-KNYSEQIIKEIK--DAHNTFTSOA 1566
Qy 959 -----LNSAPFIQSLIEVNTTKESLSNLSVAMKVQVYAQLFSTGLNTITD-----1003
Db 1567 DKSEKMMNEIKNEQIRIIDEVAKNNKSNKAI-LDIQLSVPEPKIKFKIKOLRTKSDCL 1625
Qy 1004 -----ASKVVELVSTALDETIDLTPLSEGLPIIATIIDGVSLGAAIKELSETND 1053
Db 1626 KETKDIEKISNLSIDTQETKLIENKNILNLTLEKLESKN--OKKNIEDQKKELDEVNS 1683
Qy 1054 PLLROEIEAKGIMAVNLTAATAIIVTSALGIASGFSILLVPLAGISAGISLVNNELIL 1113
Db 1684 KI--KNIESNVNQHKNY-----EIGIVEKI-----NEIA-1711
Qy 1114 QDKATKVIDYFKHISLAETEGFTLLDDKIMPODDLVLSEIDFNNSITILGKCBIRAE 1173
Db 1712 --KANK-----DOESTOKLIIITIKNLISPPKANDLEGIDTWN---LCK---YNT 1756
Qy 1174 GSGH-----TLTDDIDHFFSPS---ITVRKPWLSIYDVNLNKEKIDFSKDLMLVLPNAP 1226
Db 1757 MNIYEEFIKSYDLITHVLETVSKEPIVTEQ-----IKNKRIITAQNELLTNKNV 1806
Qy 1227 NRFGYEGWTPGFRSLNDGTGLLDRIRDHYEG-----QFYWRY-----PAFIADA 1273
Db 1807 NKAKSY-----LDDIEANEFDRIVTHFPKNKLNLDVNDKFTNEYSKVKNKGFONISNS 1856
Qy 1274 LITKLPRYEDTVNRINLDGSTRSIVPVITTE-----OIRKNLSYSFY 1317
Db 1857 -INNWK-KSTDENLNLINLQTKENYANIVSKYYSKYAEARNIFINIPKANSLNIOIK 1914
Qy 1318 GSGG-----SYSLSLSPY---NMNIDLNLV-----ENDTWVIDVNVKNITTESDEI 1362
Db 1915 SSSGIDLKFNINIALPYLDSQKDKTLTFIPSPKTSYTYTKISDSYNTLLDLKRSQEL 1974
Qy 1363 QKGE--LIENILSKLIENKIIILNHTINFYGD--NESNRFISLTFISLEIDNIIIEIDL 1420
Db 1975 QKKEQALNLIPENRLLHDKVQATNELKDTLSDLNKKKEQILNKKLLHKSNEMLNKLSC 2034
Qy 1421 VSKSY-KILLSNCMKLIENSSDIOQKIDHIGFNGEHQKIFYSYIDNETKNGFI-DYS 1478
Db 2035 NSQNTDILESCKYDIKEKSNYKEKENLGIN-----FDVKAEEQFNNDIKOIE 2086
Qy 1479 KKEGLFTAEFNSESIRNIYMPDSNNLFIYSSKOLKDIRIINKGDVKKLLIGNIFYKDDMKV 1538
Db 2087 KLENNY-----KHSEKDNVNFSEENNILQSKKCLKELT-----NAFNAEIK-2128
Qy 1539 SLSFTIEDNTIKLVNGVLDENGVAQILKFMNNAKSA-LNTSNSLMNFLESINIKNIIFYN 1597
Db 2129 -----KIBDKIIIEKNL--INKLIETRXDKMLFTYKTL---VETLKIKITDY-2170
Qy 1598 NLDPNIEFILDTNFIISGNSISQFELICDKDKNIQPYFINPKIKETSYTLVGNRQ--N 1655
Db 2171 -----TKFTITSATKFSKFKLYIDATSNLNDIN--TLQTKYDLNQNINKHVAS 2217
Qy 1656 LIVEPSYHLDOSGNI-----SSTVINFSQKLYGIDRYNVKV 1692

```

```

Db 2218 MVADAT---NDNNNLIEKEKEATKTINNLTJ--LFTIDS--NKI 2254

Search completed: November 5, 2003, 19:31:48
Job time : 59 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:47 ; Search time 15.625 Seconds  
(without alignments)  
5116.506 Million cell updates/sec

Title: US-09-126-816b-6\_COPY\_1\_1700

Perfect score: 8677

Sequence: 1 MNLVNAQLQKQVYVVKFRQ.....YLYGIDRVYVNVKIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result: being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6733.5	77.6	2366	1 TOXB_CLODI	P18177 clostridium
2	4458	51.4	2710	1 TOXA_CLODI	P16154 clostridium
3	351	4.0	2869	1 RBP1_PLAVB	Q00798 plasmodium
4	317.5	3.7	1726	1 MSP1_PLAPC	P04934 plasmodium
5	316.5	3.6	1701	1 MSP1_PLAPF	P13819 plasmodium
6	315.5	3.6	1682	1 MSP1_PLAP3	P13598 plasmodium
7	313.5	3.6	1701	1 MSP1_PLAPF	P08569 plasmodium
8	312.5	3.6	1726	1 MSP1_PLAPF	P50495 plasmodium
9	309	3.6	1639	1 MSP1_PLAPF	P04933 plasmodium
10	306.5	3.5	2136	1 YCP2_MARPO	P09975 marchantia
11	306	3.5	1630	1 MSP1_PLAPK	P04932 plasmodium
12	305.5	3.5	1957	1 SPOF_SCHPO	Q10411 schizosacch
13	301.5	3.5	1251	1 RBP2_PLAVB	Q00799 plasmodium
14	292	3.4	2748	1 NUM1_YEAST	Q00402 saccharomyc
15	290	3.3	1875	1 MLP1_YEAST	Q02455 saccharomyc
16	289	3.3	1956	1 ATX1_PLAPF	Q04956 plasmodium
17	287	3.3	3433	1 UTRO_HUMAN	P46939 homo sapien
18	280.5	3.2	2059	1 TEGU_HSV7J	P52362 human herpes
19	278.5	3.2	6885	1 SNE2_HUMAN	Q8wxh0 homo sapien
20	274	3.2	2339	1 RCP1_PLAPF	P27625 plasmodium
21	273.5	3.2	1928	1 MY51_YEAST	P08964 saccharomyc
22	272	3.1	5171	1 BPEA_HUMAN	O94833 homo sapien
23	269.5	3.1	1892	1 Y835_RICCN	Q92hd6 rickettsia
24	269	3.1	1803	1 YJL3_YEAST	P47024 saccharomyc
25	268.5	3.1	2167	1 BEN2_YEAST	P39960 saccharomyc
26	268	3.1	2663	1 CENE_HUMAN	Q02224 homo sapien
27	267	3.1	1679	1 Y109_YEAST	P40457 saccharomyc
28	266.5	3.1	1727	1 ALM1_SCHPO	Q9utk5 schizosacch
29	262.5	3.0	4092	1 DYHC_YEAST	P36022 saccharomyc
30	261.5	3.0	1628	1 NAGH_CLOPE	P28631 clostridium
31	261.5	3.0	3210	1 CENF_HUMAN	P49454 homo sapien
32	260	3.0	2014	1 YJU7_YEAST	P39526 saccharomyc
33	259	3.0	4196	1 DYHC_SCHPO	O13290 schizosacch

## RESULT 1

ID	TOXB_CLODI	STANDARD;	PRT;	2366 AA.
AC	P18177;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Toxin B.			
GN	TOXB OR TCDB.			
OS	Clostridium difficile.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1496;			
[1]				
RP	SEQUENCE FROM N.A.			
RP	STRAIN=VPI 10463;			
RC	MEDLINE=90325540; PubMed=2374729;			
RX	Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;			
RA	"Nucleotide sequence of Clostridium difficile toxin B gene.;"			
RL	Nucleic Acids Res. 18:4004-4004(1990).			
[2]				
RP	SEQUENCE FROM N.A.			
RP	STRAIN=VPI 10463;			
RC	von Eichel-Streiber C.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBSJ databases.			
[3]				
RP	SEQUENCE OF 1271-2366 FROM N.A.			
RP	STRAIN=VPI 10463;			
RX	MEDLINE=92293124; PubMed=1603068;			
RA	Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J.,			
RA	Sauerborn M.;			
RT	"Comparative sequence analysis of the Clostridium difficile toxins A			
RL	and B.;"			
RL	Mol. Gen. Genet. 233:260-268(1992).			
CC	-1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN ENTEROTOXIN			
CC	CAUSED A AND CYTOTOXIN B.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X53138; CAA37298.1; -			
DR	EMBL; X92982; CAA63562.1; -			
DR	EMBL; X60984; CAA43299.1; -			
DR	PIR; A27636; A27636.			
DR	PIR; S10317; S10317.			
DR	InterPro; IPR002479; CW binding.			
DR	Pfam; PF01473; CW binding_1; 17.			
DR	Pfam; PF04488; Gly_transf_sug; 1.			
KW	Toxin.			
SQ	SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;			

## ALIGNMENTS

34	257.5	3.0	1162	1	EXEN_CLOBO	P46082 clostridium
35	257.5	3.0	3135	1	S230_PLAFO	Q08372 plasmodium
36	255.5	2.9	2875	1	RRPL_TSWV1	P28976 tomato spot
37	255	2.9	1790	1	USO1_YEAST	P25386 saccharomyc
38	254	2.9	3144	1	VF13_YEAST	Q07878 saccharomyc
39	253	2.9	1295	1	BXAL_CLOBO	P10845 clostridium
40	251.5	2.9	8545	1	ANCL1_CAEEL	Q9n4m4 caenorhabdi
41	250.5	2.9	3305	1	APLP_MANSE	Q25490 manduca sex
42	250	2.9	4563	1	APB_HUMAN	P04114 homo sapien
43	249	2.9	1274	1	BYF_CLOBO	P30996 clostridium
44	248	2.9	1276	1	BYD_CLOBO	P19321 clostridium
45	248	2.9	3259	1	GIAN_HUMAN	Q14789 homo sapien



RA Wilkins T.W., Johnson J.L.;  
 RT "Molecular characterization of the Clostridium difficile toxin A  
 RL gene.";  
 RL Infect. Immun. 58:480-488(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN:VPI 10463;  
 RA von Eichel-Streiber C.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 CC 1- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA  
 CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE  
 CC DIFFERENT OLGOPETIDES.  
 CC DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN  
 CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL  
 CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE  
 CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X51797; CAA36094.1; -;  
 DR EMBL; M30307; AAA23283.1; -;  
 DR EMBL; X92982; CAA63564.1; -;  
 DR InterPro; IPR002479; CW\_binding.  
 DR Pfam; PF01473; CW\_binding\_1; 28.  
 DR Pfam; PF04488; Gly\_transf\_aug; 1.  
 KW Toxin; Enterotoxin.  
 SQ SEQUENCE 2710 AA; 308052 MW; 0A6B52CE84C14421 CRC64;

Query Match 51.4%; Score 4458; DB 1; Length 2710;  
 Best Local Similarity 50.7%; Pred. No. 3.7e-162;  
 Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY 1 MNLVKAQOLQKVVYKFRIOEDYVAILNALREYHNMSSVVEKYKLKDLNLTNDYL 60  
 DB 1 MSLISKEELIKUAY-SIRPRENEYKITLNLDEYNKLTNNENKYLQKLKUNESIDVPM 59  
 QY 61 NTYKSGRNKALKKPKFVLTMEVLKNNSLTPVEKNLHFVIGQINDTAINYNQWKD 120  
 DB 60 NKYKTSRRNALSNLKKOILKEVILLKNSNTSPVEKNLHFVIGEVSDIALEYIKQWAD 119  
 QY 121 VNSDYTVKPVYDSNAFLINTLKKTVESATNTNLTSEFRENLDNPDYNNKFRKRMETIY 180  
 DB 120 INAEYNIKLYWDEAPLVNLTAKAIVESSTTEALQLEBEIQNPQDFNMKPKYKRMETIY 179  
 QY 181 DKQKHFIDYKSOIBENPFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITANNNDI 240  
 DB 180 DRQKRFPIYKSOINKPTVPTIDDIKSHLVSEYNRDETIVLSSYRTNSLRKINSNHGIDI 239  
 QY 241 RNLKFAEDLVRNLQELVERWNLAASDIILRISMLKEDGGVYLDVDPILPGIQDLPKFS 300  
 DB 240 RANSLFTEGELLNYSQELLNAGNLAAASDIYVLLALKNFGVYLDVDMPLGHSDFKT 299  
 QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQRFESALSSEKSKSEIF 360  
 DB 300 ISRPSIGLDRWEMIKLEAIMKYKYYINNTYSENFDFKLDQQLKDFKLIIESKSEIF 359  
 QY 361 LPDDIKVSLPVKTAFAANNVINQALISLKDSCYGLVINOIKRYKITLNDNLNPSINE 420  
 DB 360 SKLENLNVSDLKIFAFALGVSINQALISKQGSYLTNLVIEQVKRYQLNQLHAPAIAS 419  
 QY 421 GTDFNTNTKIFSDKLASINEDNMFMKITYNLKVGFAPOVRSTINSLGPGVYTGAYOD 480  
 DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISLGPAYASAYD 479  
 QY 481 LLMFKDNSTNIHLLBEPANFPFKTKISQLTEQEIITSLWSFNQARAKSQFEYKKGYPE 540  
 DB 480 FINLQENTIEKTKASADLTIEFKFPENNLISQLTEQEIINSLWSFDQASAKYQFEKYVRDYG 539

QY 541 GALGEDNDLPAQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKISYEASCNL 597  
 DB 540 GSLSEDNGVDNKNATLADKNYLLANNKIPSNVVEEAGSKYVHYIIQLQGDISIYEATCNL 599  
 QY 598 FSKOPYSILYQKNI EGSETAYYYVAD--ABEIKDKYRIPYQISNRKNIKLTFIGHGK 655  
 DB 600 FSKNPKNSIIIQRM--NESAKSYFLSDGSGSILELNKYRIPERLKNKEKVKVTFIGHGK 657  
 QY 656 SEENTDTFANLDVSLSEIETILNLAADISPKYIEINLLGCNMFYSIVAEETYPGKL 715  
 DB 658 DEFNTSEFARLSDVSLSEISFDTIKLIDISPKNVEVNLGCNMFYSYDFNVEETYPGKL 717  
 QY 716 LLKIKDRVSELMPSISQDSITVSANQYEVINEEGREILDSHSGKWKINKERSIIDKSSK 775  
 DB 718 LLSINDKITSITLPDVKNKNSIIGANQYEVINSEGRKELLASHGKWKINKERAINSDLSK 777  
 QY 776 EYIFNPKNKIIVKSKYHLSTLQLOIRNANSSDIIDLEKKVMLTCEINVANINRQ 835  
 DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSDTKFTLNNLKLNISSIGDY 837  
 QY 836 IVEGRIEAKNLTSDSINYKNEFKLIESISDSLYDLAKHQCLDDSHSFISPEDISKTEG 895  
 DB 838 IYEEKLEPVKNI IHNISIDDLIDEFNLENVSDLEYELKLANLDEKYLISFEDISKNNST 897  
 QY 896 FRIRPINKETGNSIFETEKEIFSGYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHE 955  
 DB 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNQLDHTSQ 957  
 QY 956 VNTLSAFFIOSLIEYNTTKESLNSLVAMQVYVQALFSTGLNTITDASKVVELVSTAL 1015  
 DB 958 VNTLNAAFFIOSLIDYSSNKDVLDLSTSVKYLQALFSTGLNTIYDSIQVLNLSNAV 1017  
 QY 1016 DETDLLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLARQETEAIGTMAVNLTAAS 1075  
 DB 1018 NDTINVLTPIITTEGPIVSTIILDINGLGAAILKELDEHDPLEKUEAKGVLAINMSUSI 1077  
 QY 1076 TAITVSALGIASGFSILLVPLAGISAGIPSLVNNELIQLDKATKVIDYFKHISLAETGA 1135  
 DB 1078 AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELIHLDRATSVVNFHLSSEKGYCP 1137  
 QY 1136 FTLDDKIIMQDDLVLSEIDFNNNNSITLKGCEIWRAGGGSHLTDDIDHFFSPSITY 1195  
 DB 1138 LKTEDDKILVPTDVLVISEIDFNNNNSIKLGTNCLILAMEGGSGHTVTGNIDHFFSPSIS 1197  
 QY 1196 RKPWLSIYDVNLNIKEKIDFSLKOLMVLNAPNRPYEGYMGWTPGPRSLDNDGTKLLDR 1255  
 DB 1198 HIPSLSIYSAIGIETENLDFSKINMLPNAPSRVFWWETGAVPGLSLENDGTRLLDSIR 1257  
 QY 1256 DHYEQFYRYFAFTADALITKLPRYEDTNNRINLNGNTRSFIVPVITTEQIRKNLSYS 1315  
 DB 1258 DLYPGKFWRYFAFF-DYAITTLKPEVEDTNIKIKLDKDRNFIMPTITTNIRKNLSYS 1316  
 QY 1316 FYGCGSYLSLSPYNNMIDNLNVENDTWIDVNVVKNITIESDEIOKGLIENILSKL 1375  
 DB 1317 FDGAGGYTSLLSYSPITNINLSKDLWIFNIDNEVBESIENTGTIKGKLKIDVLSKI 1376  
 QY 1376 NIEDNKIILNHTINFYGINESNRPISLTPTSILEDINIIEIDLVSYSKYLLSGCNCK 1435  
 DB 1377 DINKNKLIGQITDFSGDIDNKKORYIFLTCELDDKISLIIEINLVAKSYLSLLSGDKRY 1436  
 QY 1436 LIENSSDIQQKIDHIGFNGEHQKIFYFVID-NETKYNGFDYDYSKKEGLFTAEFNSNII 1494  
 DB 1437 LISNLSNTEIKINTLGLD---SKNAYNYTDESNNKYFGAI-----SKTSQKSI 1483  
 QY 1495 RNINPMDNNL-----FIYSSKDL--KOIRINKGDVKLLIGNYFKD---DMKVSUS 1541  
 DB 1484 H--YKKDSKNILEFYNDSTLEFNSKDFIAEDINVEFKODINTITGKYVYDNNNTKSIDPS 1541  
 QY 1542 FTIETNTIKLVGVLDENGVAQIILKPMNNAKSALNTSNLMNPLESINKIIFYNNLDP 1601  
 DB 1542 ISLVSKNQVKNGLYNESVSYSLDFVKNSDGHENTSNMNLFLDNISFWKLFGFE--- 1598

QY 1602 NIEFIDTWNFLIGSNSIGOFELICDOKNIOPIYFINKIKETSIVLYVGNRONLIVRPS 1661  
Db 1599 NINVIDKIFTVGTNLTGVFEICDNNKNDIYFGEWKTSSKSTIFSGNGRNVVVEPI 1658  
QY 1662 YHLDGSGNISVINFSQKYLGYDIDRYVKNVXIIAPNLYT 1700  
Db 1659 YN-PDTGEDISLDSFSEYPLGYDIDRYNKVLIAPDLYT 1696

RESULT 3  
RBPI PLAVB  
ID RBPI PLAVB STANDARD; PRT; 2869 AA.  
AC Q00758;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Reticulocyte binding protein 1 precursor.  
GN RBPI.  
OS Plasmodium vivax (strain Belem).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92315338; PubMed=1617731;  
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax  
merozoites";  
RL Cell 69:1213-1226(1992).  
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
CC HUMAN RETICULOCYTE CELLS.  
CC -!- SUBUNIT: Homodimer (Potential).  
CC -!- SUBCELLULAR LOCATION: Membrane-bound.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M8097; AAA29743.1; -  
KW Malaria; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.  
FT DOMAIN 18 2807 EXTRACELLULAR.  
FT TRANSMEM 2808 2826 POTENTIAL.  
FT DOMAIN 2827 2869 POTENTIAL.  
FT SITE 1030 1032 CYTOPLASMIC.  
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 2869 AA; 330213 MW; 89DBE442205EBCFF CRC64;

Query Match  
Best Local Similarity 19.7%; Score 351; DB 1; Length 2869;  
Matches 399; Conservative 323; Mismatches 633; Indels 626; Gaps 105;

QY 6 KQALQKQWVVKFRIQDEY-----VAILNALEEYHNMSESSVVEKYLK-LDNNL- 55  
Db 946 KALKKIVSDSLRDKIDQYETEFKETSVAENTVSTIQSLSKA--IDSLKRLNGSINCK 1003  
QY 56 ---TD-NYLTNLYKSGRNKALK-----KFKEYLTMEVLELKNLSLTPEVKNLFIWIG 104  
Db 1004 KYNTDILRLSKIKTLREVEQKMPKRGDKCGENTALLKLSLRDMKMGKINEKLN-----D 1059  
QY 105 GOIN--DT-----AINYINQKQV-----NSDYTV- 127  
Db 1060 GRNLSLDTKKEDLLKFYSSESKSIHLSKDQKQDPQLNRIDSEWDIKRDVDELNVYQVI 1119  
QY 128 -----KFVYDGNFL-----INTLKTIVESATNTLESFRE-----NLNDPEFDYK 170  
Db 1120 SENKVTLPKNYSVTYIEAMHSHINTVAHGIT-SNKNEILKSVEKDEKLNLVQEQNEDYK 1178  
QY 171 F---YRKRMEII---YDQKHFIDYKSGIENPEFIIDNIKTVLSN-----EYSKDLE 219

Db 1179 VKNPENKQLEAIRGSMKSLKEVINKHVSEMTQ-----LESTANTLKSNAKGNEHDL 1233  
QY 220 ALNKY-----IESLANKI-----TANNNDIRNLEKFADEDLVLYNQELVERWNLA 266  
Db 1234 ELNKTGQMRDIYEKLAKIABELKEGTWNLKDANERKANKVEPERNIIGHVLE----- 1288  
QY 267 AASDILRISMLKEGGVYLDVLDILPGIQDLFKSINKPDSITNTSWEMIKLEAMKYKEY 326  
Db 1289 -----RITVEKQAGV--VEEMNSLKTIEKIQE-----TSDDSQNELVTTSTIKHLE 1337  
QY 327 IPGYTSKNFMDLREVOVSRESALSSKDKSEIEFLPLDDIKVSPLEVKI-----AFANNSV 382  
Db 1338 AKGY-----EDVIKNEEDSIQLR-EKAKSLETLEDEMKKLVQVNNMIQSAIQGNAG 1388  
QY 383 INQALISLSDSYCDLVINQIKRYKILNDNLNPSINEGTDFNTMTKIFSKLA----- 436  
Db 1389 ISKEL-----NELKGVIELLISTNYSILLEYVKKNSSSVRFPSQLANGEFTK 1435  
QY 437 SISHEDNMFMKITYNLVKVGFAPDVRSTINLSGFGVYTGAYQDLLMFKDNSTNIHLEP 496  
Db 1436 AEGEKNASARLAEAEKLEQIVKDL-----DYSD-----IDD 1469  
QY 497 ELRNFEPKTKISQLTEOITSLMSFNOARAKSOFEEYKKGYPFGALGEDDNLDAQN-- 554  
Db 1469 KVKIEGKREILKMKESALT-FWESE-----KFKQMCSSHMEKAKGKKKIEYKONG 1522  
QY 555 -----TVLDK-----DYVSKILSMKTRNKRYIHYIQOGDKISYEASC----- 595  
Db 1523 DGKANITDSQMEEVGNVSK-----AEHAFHTVEAQVDKT--KAFCEISIVAVYT 1570  
QY 596 ---NLFSKDPYSSILYQKNIEGSETAYYYYVADAEIKEDIKYRIPYQIS-NKRN- 646  
Db 1571 KQDNLFNESLMKVEKVKCEKKNDEAEKY-----SAKLKPYDG-RIKARVSENERKISELKE 1625  
QY 647 KLTFIGHGKSFN--TDFFANLDVDSLSEIETILNLAADISPKYIEINLGCNMFYSYI 705  
Db 1626 KAKVEKESOLNDVSTYKSLQIDNCRQLDSVL-----NIGRVKQNALQYFD 1674  
QY 706 YAEITYPKLLKTKIDRVSELMPSISQDSITVSANQYEVRIENERGKREILDHSGKWINK 765  
Db 1675 SADKSMKSVL-----PISLGAESKSLDKVAAKESYEKNL-----ETVQNMESRINVE 1722  
QY 766 ESIILKIDISSKEYISFNPKENKIIIVKSKYLHELSTLQEIERNNANSDDIDLEKKVMTCE 825  
Db 1723 EGSULTDIDKK-----ITDIENDLLKMKQYEE--GLLQKIKENADRKSNF-----LVGSE 1772  
QY 826 INVASNIDRQI-VEGRIEEAKNLTSDSINY-----IKNEPKLIESDSLYDLKHQNGL 878  
Db 1773 INALLDPSTSIPIKLKKE-YDMTGDLKQYGVKQNEIHGEF-----TKSYNLIETHLSNAT 1827  
QY 879 DDSHFISPE-----DISKTENG-----PRIRPINKETGNSIFITEKEIFSEYA 922  
Db 1828 DYS--VTPEKAQSLRELAEEKEEHLRRREBEAIFLLANDIKKVESLKLKEMMKVSAEYE 1885  
QY 923 -----THISKEISNIKDTIEDNVNG-----KLVKCV-----NLDAHEVN 957  
Db 1886 GKKRDHTSVSOLQDMK-TIVDELKTLNDISECSSVLNNVSVIVKVKESKHADYRRDAN 1944  
QY 958 -----TLNSAFF-----IOSLIEYNTTKESLNSLVAMKVQVYAOLFSTGLNTITDAS 1005  
Db 1945 SWYESWVTLANYFLSDEAKISSGMFNA--EMKSNFKTDLEIF-----SVISNSN 1994  
QY 1006 KVELVSTALDETDLPTLSEGLPIIATIDGVSLSGAAIKELSETNDPLRQIEIAKIG 1065  
Db 1995 ELAKKIEQDSNDVIQ-KERESEQLAKDATIYV-----IKLKNFEKEL--EEAKNKE 2046  
QY 1066 IMAVNLTAASATVTSALGIASGSILLVPLAGISAGIPSLVNNNELIILQDKATKVIDYFK 1125  
Db 2047 VVSEKVRALKRL-SQVEGIRCHFE-----NPHRLDN-----TELENLK 2086  
QY 1126 HISLAETEGATLLDDKLIIMPQDDLVLSEIDFN--NNSITL-----CKCEIMRAE 1173



Db 2087 KM-----VTIYDKKSERESGLQEMENMTYSNSITQLEGIVSAGESKEDIKLE 2138  
 QY 1174 GSGH-----TLTDIDHFRSSPSITVR-----KPMLSIYD-VLNKKKIDISK 1217  
 Db 2139 RSNEWRNISEKISTIDSQVEMWSTIDELYKGNQAHWLSLSYANMKTSK---K 2194  
 QY 1218 DLWVLNAPNRVFGYEMGWTFGRSLDNDGTLKLDRIRDHYBGQFYWRFYAFIADALITK 1277  
 Db 2195 LIMINKENT-----EKVDYIKDNSS-----TDGYVET 2225  
 QY 1278 LKPRYEDTVNRNLGNTRSFUPVITTEQIRKNLSYSPYGGSGYSLSLSPYNNIDLN 1337  
 Db 2226 LK-----GPIYGLTFSSA-----SE 2241  
 QY 1338 LVEN-DTWIDVDNVVNKTIIESDIOXGELIENILSKLNIEDNKIILNHNTHINFGDIN 1396  
 Db 2242 IVQADTVSNVPAKHEKESLNARIDIKELYLFHONSISIVEGGV---QNMALYDKLN 2298  
 QY 1397 ESNRFISLTSILEDINIIIEIDLVSYSKILLSGCMKLIENSDDI-QOKID-HIGFNG 1454  
 Db 2299 BEKR-----EMDELVRN---ISETKLKQMEHSTDVFKPMIELHKGWNE 2338  
 QY 1455 EHQKIFYSIDNETYNGFDISK-----KEGLFTAEPFNSIIRNIYMPDNNULFI 1507  
 Db 2339 TNKK---SLEKEKKLSVNDHMSMEAEMIKNGL---KYTPES-VQNI-----NNIYS 2385  
 QY 1508 YSSDKLDKIRIINK---GVKLLIGNY-FKDDMKVSLSTIEDTNTIKLNGVYLDENGVAQ 1564  
 Db 2386 VIEAEVTKLEEIDRDYGD-----NYQIVEEHKQFSLIDRTNAL-----MDD---IE 2430  
 QY 1565 ILKFMNNAKSALNTSNLMNPLESINIKNIFFYNNLDPIETLDTN-FIIGSNSIGO-- 1621  
 Db 2431 IFKKN-----YLMEVATETIHRVNDYIEKTNKLVQAK 2466  
 QY 1622 --FELICD-----KKNQIOPYFINKIKETSY--TLVGNRQNLIVPEPSYH 1663  
 Db 2467 TEYEQILENIKQNDMLQNIPLKVSITIEYFENVKKESILNDLY---EQERLLKIGEH 2523  
 QY 1664 LDD-SGNISSTVINP--SQKY-----LYGIDRYNKKV 1692  
 Db 2524 LDEIKRNVTTLSYIEIDQKEMMSKNLLEKKSQNNYTSIYELBREANEI 2574  
 RESULT 4  
 MSPL\_PLAFC  
 ID MSPL\_PLAFC STANDARD; PRT; 1726 AA.  
 AC P04934;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
 DE (PMNSA) (P195).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate Camp / Malaysia).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=5835;  
 RN [1]  
 RP SEQUENCE OF 1-1103 FROM N.A.  
 RX MEDLINE=86205236; PubMed=3517809;  
 RA Weber J.L., Leininger W.M., Lyon J.A.;  
 RT "Variation in the gene encoding a major merozoite surface antigen of  
 the human malaria parasite Plasmodium falciparum."  
 RL Nucleic Acids Res. 14:3311-3323(1986).  
 RN [2]  
 RP SEQUENCE OF 1104-1726 FROM N.A.  
 RX MEDLINE=88143999; PubMed=3278296;  
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;  
 RT "Merozoite surface protein sequence from the Camp strain of the human  
 malaria parasite Plasmodium falciparum."  
 RL Nucleic Acids Res. 16:1206-1206(1988).  
 CC -1- SURCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 (Potential).  
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42

kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X03831; CAA27446.1; -.  
 DR PIR; A23386; SAZQM.  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00008; EGF; 1.  
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 FT Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.  
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1726 AA; 196197 MW; D88AD45FA3528CF3 CRC64;  
 Query Match 3.7%; Score 317.5; DB 1; Length 1726;  
 Best Local Similarity 19.9%; Pred. No. 6.2e-05;  
 Matches 350; Conservative 284; Mismatches 602; Indels 523; Gaps 89;  
 QY 51 DINLNTDNYLNTYKSGENKALKKFEVLTWEVLEKNSLTPVEKNLHFWIGQINDY 110  
 Db 154 DLKRVRYNLT-----IKELKYPFLDLTNHMLTCLD-NTH----- 189  
 QY 111 AINY-INQKDVNSDYTVKFDVDSNAFLINTLKTIVESATNNTLE-SFRENLDPEPDY 168  
 Db 190 GPKVLIDGYEIN-----ELLYKLN-FYFDLLRAKLVNDVANDYCOIFNLKIRANELDV 243  
 QY 169 NK----FYRKMEIYDKQHFIDYKSOIENPEFIIDNTIKYLSNEYSKOLBALNKY 224  
 Db 244 LKLVFGYRKPLDNIKNVGMEDYIKKN-----KTTIAN-----INEL 282  
 QY 225 IEESLTKTANNNDIRNLEK---FADEDIYRLYNQELVERNWLAAASDILRISMLKEDG 281  
 Db 283 IEGSKTTIDQKNADNEEGKKLYQAQYDL-SIYNKQLEEAHNLISVLE-KRIDTLKKN 340  
 QY 282 GVLVDVILPGIO-PDLFKSINKPDSINTSWEMIKLEIMKYKEYIPGYTSKNFMDLDE 340  
 Db 341 NIKELDKINEIKNPPANSNGTNTLDDKNKKEEHEE--KIKE-IAKTIKFNIDSFT 397  
 QY 341 EVQSFESALSSKSKSEIFLPLDDIKVSPLEVKIAFANNVINOALISLKSDSYCSDLVI 400  
 Db 398 D-PLELEYLRKKNKVDVTPKSDPTKSVQIPKVPYPNGIVYVPLPLTDIHNSLAAD--- 453  
 QY 401 NQIKNRY-KILNDNLNPSINEGTDFNTWKIPSDKLASINEDNMFMFKITNYLKVGFA 459  
 Db 454 -NDKNSYGLMNPDPTEKINE-----KIITD-----NKRKIFT-----NNIK- 490  
 QY 460 PDVRSTINLSGPGVVTGAYQDLLMFKDNSTNIHLEPELRFNPFEPFKTKISQITQSITSL 519  
 Db 491 -----KQIDLEEKINHTEKQNKLL 511  
 QY 520 WSNQARAKSQFEYKKGYPFEGALGEDNDNLDFQAQTVLDDKDYVSKILSSMKTRNKE--- 576  
 Db 512 EDYE--KSKDYEELEKFEY-----MKFNNN--FDKDVVD-KIFSARYTYNVEKOR 558



QY 1004 -----ASKVVE-LYSTALDETIDLLPTLSEGLPIIATIIDGVS----- 1041  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 999 IKQFVKNSKVIITGTETQKVNLDEIKKLKDYLQISFDLYNKYKULKDLFLFNKKELQG 1058  
 QY 1042 -GAAIKELSETNDPLRQEIRAKIGIMAVNLTAASTAIVTSALGIASGFSL--VPLAG 1098  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1059 DKMQIKKLT-----LLKEQLSKLSL-----NNPHNVLPNQSPFFNKKEAE 1101  
 QY ISAGISLVNNELLQDKATKVIDYPKHISLASTEGATLLDDKIIMPQDDL-----V 1151  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1102 IASTENTLENTKLLXHYKGLVKYNGESPLKT-----LSEVSIQTEDNYANLEKFRA 1155  
 QY LSEIDFN-NNSTILGKEIWRABEGSGHTLTDIDHFFSSPSITYRKPMLSIYDVLNIKK 1210  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1156 LSKIDGKLDNLHLGKKKSFLSGLHLITE-----LK 1189  
 QY EKIDFSKOLMVLNAPNRVFGEYMGWTPGFRSLDNQTKLLDIRDHYGQFYRYFAFI 1270  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1190 EVI---KNKNYTGNP-----SENKKVNEALKSYEN-----FL 1220  
 QY ADALITYKL----KPRVEDTNVRINLGNTSRFPVPITTEQIRKNI.SYSFGYGGGSYSLS 1326  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1221 PEAKVTVVTVPPQPDPVTPSPLSRVVSGSGS-----TKEETQIPTSGSLL-TELQQVVQ 1273  
 QY LSPVNMIDLNLV-----ENDTWV-----IDVDNVVNKITTESDEIQ----- 1363  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1274 LQNYDEDDSLVLPFGSESDNDELQDVQVTEAISVTMDNLISFENEYDIVYUKPLA 1333  
 QY -----KGELIENILS-KLNIEDNKII.LNNH--TINPYGDINESN----RFISLTFSILE 1410  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1334 GYVRSCLKQIEKNIITFNLNLD---ILNSRLKRRKYFLDVLLESSDLMQFKHISSNEVII 1390  
 QY DINIIEIDLVSXSXYKILLSGCNKMLESDDIOQKIDH-IGNGSHQKY--IFYSYID 1466  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1391 D-----SFKLLNSEQNTHLLSKYKIKESVENDIKFAQEGISYEKVLAKYKD 1438  
 QY NETKYNQFI-----DYSKKEGLFTAEBFSNESIIRNIYMPSDNL 1505  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1439 DLESIKVKIKEEKEFPSPPTPPSPAKTDEOKESKFLPUTN---JETLY-----NNL 1491  
 QY FIYSSKDLDIRINKGD-----VKLLIGNYFKDDMKVLSFTTIEDTNTI--KLVGVY 1556  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1492 -----VNKIDDYLNLKAKINDCNVEKDEAHVKIT-KLSDLKAIDDKID-LF 1536  
 QY LDENGVAQILKFMNAKSA.MNTSLMNFLESINIKNIFNNL 1599  
 ||| : | : | : | : | : | : | : | : | : | : | : | : | :  
 DB 1537 KNTNDFEAIKGLIND-----DTKOMGLKLLSTGLVQNFPNTI 1574

RESULT 6  
 MSP1\_PLAF3 STANDARD; PRT; 1682 AA.  
 ID MSP1\_PLAF3 AC P19598; Q25921;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMMSA) (p190).  
 DE DE MSP-1.  
 GN Plasmidium falciparum (isolate ro-33 / Ghana).  
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OC NCBI\_TaxId=5834;  
 RN [1].  
 RP SEQUENCE OF 1-1061 FROM N.A.  
 RX MEDLINE=88166657; PubMed=3327688;  
 RT Certa U., Rotmann D., Matile H., Reber-Liske R.;  
 RL "A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats."; EMBO J. 6:4137-4142(1987).  
 RL [2].  
 RP SEQUENCE OF 1032-1682 FROM N.A.  
 RX MEDLINE=95354793; PubMed=7628566;

RA Tolle R., Bujard H., Cooper J.A.;  
 RT "Plasmodium falciparum: variations within the C-terminal region of  
 RL mezoite surface antigen-1";  
 RL Exp. Parasitol. 81:47-54 (1995).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC (potential).  
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 CC kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M35727; AAA29715.1; -  
 DR EMBL; Y00087; CAA68280.1; -  
 DR EMBL; Z35326; CAA84555.1; -  
 DR InterPro: IPR006209; EGF\_like.  
 DR Pfam: PF00008; EGF; 1.  
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1682  
 FT TRANSMEM 1666 1682  
 FT CARBOHYD 233 233  
 FT CARBOHYD 462 462  
 FT CARBOHYD 528 528  
 FT CARBOHYD 599 599  
 FT CARBOHYD 785 785  
 FT CARBOHYD 801 801  
 FT CARBOHYD 901 901  
 FT CARBOHYD 947 947  
 FT CARBOHYD 1071 1071  
 FT CARBOHYD 1178 1178  
 FT CARBOHYD 1569 1569  
 FT SEQUENCE 1682 AA; 192462 MW; C82A1E159948AD6 CRC64;

Query Match 3.6%; Score 315.5; DB 1; Length 1682;  
 Best Local Similarity 19.5%; Pred. No. 7.1e-05;  
 Matches 334; Conservative 281; Mismatches 585; Indels 511; Gaps 83;  
 QY 36 NMSESSVVEKYLKLDINLTDNYLTNYKSGRNKALKKFEYLTMVLELKNLSLTPE 95  
 DB 100 NPSDDSSDSDAKSYADLKHVRQNYLFT-----IKELYPELFDLTNHLTLCD 147  
 QY 96 KNLHFIWIGQINDTAIY- INQKDVNSDYTVKPYDSDNAPLINTLTKTIVESATNNTL 154  
 DB 148 -NIH-----GFKYLDIGYEIN-----ELLYKLN-FYPDLRAKLVDCANDYC 189  
 QY 155 E-SFRENLPDPFNK-----PYRKRMEIYDKQHFIDYKYSQTEENPEFTIDNIITY 209  
 DB 190 QIPFNLKIRANLDELVKLVFGYRPLDFIKDQVGMEDYIKKN-----KTT 236  
 QY 210 LSNEYSKOLEANKYIESLNKITANGNDIRNLEK---FADELVRLYNQELVERWILA 266  
 DB 237 TAN-----INELIEGSKTIDQKNADNEEGKKLYQAYDLF- IYNKQLQBAHLI 287  
 QY 267 AASDILRISMLKGDGVVDVILPGIOPDLFKSINKPDSINTSWEMIKLEAIWKYKE 326  
 DB 268 SVLE-KRIDTLKKNENIKKLEDDIKDKIDAEKPTGVNQILS-----LRLEKSRHEEK 341  
 QY 327 IP--GYTSK-NPDMLEDEVQSFESALSSKSDKSEIFLPDDIKVSPLEVKIAFANNSVI 383  
 DB 342 IKEIARTIKFNIDRLFTD-PLFLEYLREKKNKVDVTPKSQDPKTSQVQIPKVPYNGIVY 400  
 QY 384 NQALISLSDSYCDLVINQIKRY-KILNDNLNPSINEGTDENTMTKIFSDKLASINED 442  
 DB 401 PLPLTDIHNLSAAD-----NDRKSYGDLNMPHTKEKIN\*\*\*\*\*KIITD-----NKE 442

QY 443 NMFMKITNYLKVGFAPDVRSTINLSGPGVYVTGAYQDILLMFKDNSTNIHLLEPELRNFE 502  
 DB 443 RKIFI-----NNIKQIDL-----EKKIN 462  
 QY 503 FPKTKISOLTEQITSLWSFNQARAKSQFEBYKKGYPFEGALGEDDNLDPAQNTVLDKDYV 562  
 DB 463 HTKEQNKKLE-----DYKSKDYEBELLEKFYEMKFNPNFKD-----VVDKIFS 508  
 QY 563 SKKILLSMKTR-NKEYTH-----YIVQLOGDKISYEASCNLFPSDPYSSILYQKNIESE 616  
 DB 509 ARYTNYVEKQRYNNKFNSSNNNSVYNQVKLKALSY-----LEDYS---LRKGISBKD 557  
 QY 617 TAYTYVY---ADAETKEIDKYRIPQVSNKRN--IKLTFIG--HGKSEFNTDTFANLDV- 668  
 DB 558 FNHYTYLKTGLEADIKKLE-----EIKSSENKILEKNFKGLTHSAN-----ASLEVS 605  
 QY 669 DLSSEIETILNLAKADISPKYIEINL---LGCNMFYSYSIYAEETYPGK---LLLKIK 720  
 DB 606 DIVKLQVKVLLIKKIEDLRK-IELFLKNAQKDSIHVPNIYKPNKPEPYVLIVLKEV 664  
 QY 721 DRVSELMF-----SISQDSITYSANOYE-----VRINEGKEIILDSHGW 761  
 DB 665 DKLKEFIPKVKMDLKKQAVLSSITQPLVAASETTEDGGHSTHTLSQSGETEVTETSET 724  
 QY 762 INKESIIKOISKEYISFNPKENKIIIVKSKYLHELSTLLQEIERNNANSSDIDLEKVM 821  
 DB 725 VGHITVT-----ITLPPKEVNVENS-----IEHNSNDNSQALTKTYL 764  
 QY 822 TECEINVASN-----IDRQIVEGRIEBAKNITS-----DSINYIKNEFKLI 862  
 DB 765 KKLDEFLLTKSYICHKYLIVSNSSMDQLLEVYNTLPEENELKSCDRDLLFNQNNIPAM 824  
 QY 863 ESISDSL-YDLKQHQGLDDSHFISFEDISKTEGFRFRFNKETGNSFIFETEKEI----- 917  
 DB 825 YSLYDSMNNDLOH-----LFFELYQKEYIYLLHKLKEENHIKKLLEPKQITGTS 874  
 QY 918 -----FSEVATHIS--KEISNIKOTIFDN-----VNGKLVKKNLDAAEVNTLN 960  
 DB 875 STSPGNTVNTAQSATHSNQSNASSTNTQNGVAVSSGPAVVBESHDPPLTVLSISN 934  
 QY 961 SAFFIQSLIEY-NTTK--ESLSNLSVAMK-----VQVYAQLFSTGLNT-ITDASKVVE- 1009  
 DB 935 DLKGVSLNLGNKTKVPNPLTISTTEMEKFYENILKIMIPIDNDIKQFVKSNSKVITG 994  
 QY 1010 LVSTALDETDLTLSEGLPIATIIDGVSL-----GAAAKEISETNDP 1054  
 DB 995 LTETQKALNDEIKKLDKDTQLSPDLNKKYKLDLRFNKKKELGQDKQKQKUT----- 1049  
 QY 1055 LLROEIEAKIGIMAVNLTAASATVTSALGIASGFSIL--VPLAGISAGIPSLVNNELI 1112  
 DB 1050 LLKQLESKLSL-----NNPHNVLQNFVSFFNKKKEAEIAETENTLENTKIL 1097  
 QY 1113 LQDKATKVIDYFKHISLAETEGAPTLDDKILIMPQDDL-----VLSIEDFN--NNSITL 1164  
 DB 1098 LKHVKGLVKVYNGESSPLKT-----LSEVSIQTEDNYANLEKPRVLSIKDGLKLDNLHL 1151  
 QY 1165 GKCEIWRABEGSGHGLTDDIDHFFSSPSITYRKPLWSIYDVNLNKEKIDSKOLMVLPN 1224  
 DB 1152 GKKKLSFLSLGHLHLITE-----LKEVI-----KKNKYTGN 1182  
 QY 1225 APNRVFGYEMGWTGFRSLONDGTGKLLDRIDRHYEGQFYWRFYAFIADALITKL----- 1280  
 DB 1183 SP-----SENNKVNALKSYEN-----FLEAKVTVTVPQP 1216  
 QY 1281 RYEDTNVRINDGNTRFIVPVITTEQIRKNLSVSYFGSGSGSYSLSPYNNIDNLNV- 1339  
 DB 1217 DVTPSPLSVRVSGSGS-----TKBETQIPTSGSL--TELQVQVQLQNYDEEDSLVVL 1269  
 QY 1340 -----ENDTWV-----IDVNVVKNITIESEIQ-----KGLTENI 1371  
 DB 1270 PIFGESEDNBYLDQVVTGSAISVTMDNILSGFENFEDVYILKPLAGVYRSLKKQIEKNI 1329  
 QY 1372 LS-KLINEDNKIILNNH--TINFYGDINESN-----RFISLTFISILEDNIIEIDLVS 1424

Db 1330 FTFLNLND---ILNRLKRRKYFLDVLSDLMQFKHSSNYIED-----S 1374  
 QY 1425 KYLLSGCMKLIENSSDIQKIDH-IGFNGHQY---IFVSYDNETKNGFT----- 1475  
 Db 1375 FKLLNSEQNTLLKSYKIKESVENDIKPAQGGISYKVKLAKYKDDLESIKVKKEKE 1434  
 QY 1476 -----DYSKKEGLFTAERFSNESIIRNIYMPDSNNFIYSSKDLKDIRIIN 1520  
 Db 1435 RPSSPTTPPSAKTDEQKESKFLPFLTN---IETLY-----NNL-----VN 1474  
 QY 1521 KGVKLLIGNYFKDMKVSLSFTIEDTNTIK 1551  
 Db 1475 K-----IDYLYNLKAKINDCNVEK 1494

RESULT 7

MSPI\_PLAFM STANDARD; PRT: 1701 AA.  
 AC P08569;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
 DE (PMMSA) (P190).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=70153;  
 RN [1]  
 RN [2]  
 RN [3]  
 RA Tanabe K.;  
 RA Tanabe K., Mackay M., Goman M., Scaife J.G.;  
 RT "Allelic dimorphism in a surface antigen gene of the malaria parasite  
 Plasmodium falciparum.";  
 RL J. Mol. Biol. 195:273-287(1987).  
 RN [2]  
 RP REVISIONS TO 1403; 1569 AND 1629.  
 RA Tanabe K.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-115 FROM N.A.  
 RX MEDLINE=86136024; PubMed=3004972;  
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
 RA Stunnenberg H., Bujard H.;  
 RT "Polymorphism of the precursor for the major surface antigens of  
 Plasmodium falciparum merozoites: studies at the genetic level.";  
 RL EMBO J. 4:3823-3829(1985).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 (Potential).  
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 CC EMBL; X05624; CAA29112.1; --  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00008; EGF; 1.  
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19 POTENTIAL  
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AP96EA98 CRC64;  
 Query Match 3.6%; Score 313.5; DB 1; Length 1701;  
 Best Local Similarity 19.9%; Pred. No. 8.6e-05;  
 Matches 357; Conservative 284; Mismatches 601; Indels 549; Gaps 91;  
 QY 36 NMSESSVVEKYLKLDINNLTNTYKKSGRNALKKFKKEYLTMEVLEKNNLSLTPVE 95  
 Db 106 NPSNDSSTNTYADLKHRYQNYLFT-----IKELKYPEFLDLNRLMLT-IS 152  
 QY 96 KNLHFIWGGQINDTAINYINQMKDVNS-DYTVKFVYDSNAFLINTLKKTIIVESATNNTL 154  
 Db 153 KNV-----DGFKYLIDGYEBEINELLYKLNFFYD-----LLRAKLNDACANSVC 195  
 QY 155 E-SFRENLDPEFDYK----FYRKRMEIIVDKQGHFDYDKYSQTEENPEFIIDNIKIY 209  
 Db 196 QIPFNLIKIRANELDVLKIVFGYRKPLDNKDVGMEDYIKKN-----KTT 242  
 QY 210 LSNEYSKDLKALNKYIEESLANKITANNNDIRNLEK--FADEDLVRLYNQELVERNLA 267  
 Db 243 IAN-----INELIEGSKTIDQNKADNEEGKKLYQAYNLFIYNNKLOEAAHNLIS 294  
 QY 268 ASDILRISMLKEDGGVLDVILPOIQDPLFKSI--NKPDSITNTSWEMIKLEAIMKYKE 325  
 Db 295 VLE-KRIDTLKKNENIKKLEDDIKITDAENPTTGSFKNPLPENK-----KK 341  
 QY 326 YIPGYTSK-----NFDMLDEVOVSFESALSSKSDKSEIFPLDDIKVSPLEVKI 375  
 Db 342 EVEGHEEKIKETAKTIKENIDSLFTD-PLELEYLREKNKKVDVTPKSDQTPKSVQIPKV 400  
 QY 376 APANNSVINQALISLSDSCVSLVINOIKNRY-KILNDNLNPSINEGTDFNTMTKIFSDK 434  
 Db 401 PYPNGIVYPLPLTDIHNSLAAD-----NDKNSGDLNPNPTKEKINE-----KIITD- 447  
 QY 435 LASISNEDNMFMFIKITNLYLVGFAPDVRSTINISGGVYTGAYQDMLLMFKDNSTNIHLL 494  
 Db 448 -----NKKERKAFI-----NNIKKQIDL----- 464  
 QY 495 EPELANFPFPKTKISQLTQEITSLSWFSNQARAKSQFBEYKKGYPFEGALGEDNDLDPAN 554  
 Db 465 --EKNINHTKEQNKKLLE-----DYEKSKDYBELLEKPYE-----MKFNNN 505  
 QY 555 TVLDKDYYSKILSSMKTRNKE---YIH-----YIVQLQDKISVEASCNLFSPDKP 602  
 Db 506 --FDKVDVY-KIFSARYTYNVEKQRYNNKFSNNNSVYNVQKLKALSY-----LSD 554  
 QY 603 YSSILYQKNIEGSEYAYIVYV--ADAEIKETDKVIRIPVOISNKN--IKLTFIG--HCK 655  
 Db 555 YS---LRGISEKDFNHYTTTKGLEADIKLITE-----EIKSSENKILEKNFKGLTHSA 606  
 QY 656 SEFNTDTFANLDV-DLSLSEIETILNLAADISPKYIEINL----LGCNMFYSYSIAEET 710  
 Db 607 N-----ASLEVSIVDKVQVQVLLIKKIEDLRK-IELFLKNAQLKDSIHVINYKPN 658  
 QY 711 YPGK-----LLKIKDRVSELMPSTISQ-----DSIT---VSANQVEVRINEEG--KR 752  
 Db 659 KPEPYLLVLKKEVDKLKEFIPKVKMDLMEKQAVLSSITQPLVAASE-----TTEDGGHST 714  
 QY 753 EILDHSGKWINKEESIIMD-----ISSEKYSFNFNKNKIIVKSKYLHELSTL 800  
 Db 715 HTLSQSGTEVEETEVEETEVEETVGTHTTTITLPPKEESA--PKEVKVENS----- 763  
 QY 801 LOEIRNANSSDIDLEKKVMLTECE-----INVA-SNIDRQIVEG---RIEBA 844

```
Db 764 ---IEHKSNDNSOALTKTVLKKLDEFLTKSYIKYILVNSNMDQKLEVLNLTPEE 820
Qy 845 KNLTS---DSINVIKNEFKLIESISLSL-YDLKHQNGLLDSHFISFEDISXTENGFRIR 899
Db 821 KELKSCDPLDLLFNQNNIPAMYSULYDSMNLDQH-----LFFELYQEMIIYLH 870
Qy 900 FINKETGNSIFITEKEI-----FSEYATHIS--KEISNIKOTIFDN--- 939
Db 871 KLKEENHIKLLBEQKQITGTSSTSPGNTTNTVAQSAATHSNQCSNASSTINTQNGVA 930
Qy 940 -VNGKLVKKVNDAAHEVNTLSAPFIQSLIEY-NTTK-ESLSNLSVAMQVYQAQLFST 996
Db 931 VSSGPAVVESSHDPVLVLSNDLKGIVSLNKGKTKVPNPLTITTEMEFENILKN 990
Qy 997 GLNTITD-----ASKVVE-LVSTALDETIDLLPTLSEGLPIIATIDGVSL----- 1041
Db 991 NDTYFNDDIKQFVKSNSKVITGLTETQKVALNDEIKKLKDTQLQSPDLTKNKKLKLDRUF 1050
Qy 1042 -----GAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAIVTSALGIASGFSIL 1092
Db 1051 NKKKELGQDKWQIKKLT-----LLKEQLSKLNSL-----NNPHNVLQNFSVF 1093
Qy 1093 L--VPLAGISAGIPSLVNNELILODKATKVIDYFKHISIAETGAPTLLDDKIIMPQDDL 1150
Db 1094 FNKKKEAETAEENTLENTLKILKHKKYGLVKYNGESSPLKT-----LSEVSIQTEDNY 1147
Qy 1151 -----VLSEIDFN--NNSITLKGCEIWRAGSGGHTLDDIDHPSSPSITYRKPWLSI 1202
Db 1148 ANLEKPRALSUKIDGKUNLHGGKKLSFLSSGLHLHTE----- 1187
Qy 1203 YDVLNLIKKEKIDFSDKDLVLPNAPNRVFGYEMGWTGPRSLNDGDKLDRIRDHVEGOF 1262
Db 1188 -----LKEVI---KNKNYTGNSP-----SENKKVNEALKSVEN--- 1218
Qy 1263 YWRYPAFIADALITKL-----KPRYEDTNVRINLDGNTSRFIVPVTTEQIRKNLSVSFVG 1318
Db 1219 -----FFPEAKVTVTVTPQPDVTPSPLSVRVSGSGS-----TKERTQIPTSGSLI- 1265
Qy 1319 SGGSYLSLSPPYNNMIDNLV-----ENDTW-----IDVDNVVKNITIEDS 1360
Db 1266 TELQVQVQLQNTDEEDDSLVVLPFQESDNDDEYLDQVVTGSAISVTMDNLSGFEYND 1325
Qy 1361 EIQ-----KGELIENLS-KLNTIEDNKILNNH--TINFYGDINESN-----RPI 1402
Db 1326 VIYKPLAGVYRSLKQIEKNIITFNLD--ILNSRLKRRKYFLDVLSDLMQPKHI 1382
Qy 1403 SLTFSILEDINIIEIDLVSQKILLSNCWKILNSDSIOQKIDH-IGFNGEHQKY-- 1459
Db 1383 SSNEYIIE-----SFKLNSRQKNTLLKSYIKESVENDIKFAQEGISYVE 1430
Qy 1460 -IFSYIDNETKNGFI-----DYSKKEGLFTAEPFNSIIRNI 1497
Db 1431 KVLAKYKDDLESIKKVIKEKEKFPSPPTTPPSAKTDQKESKFLPFLTN---IETL 1487
Qy 1498 YMPDSNNLFYSKDLKIRINKGD-----VKLLIGNYFKDDMKVSLSTFIETDNTI 1550
Db 1488 Y-----NNL-----VNKIDDYILNLKAKINDCNVEKDEAHVKIT-KLSDLKAI 1529
Qy 1551 --KLNGVYLDNGVAQILKPMNAKSALNTSMLNLFESINIKNIFYNL 1599
Db 1530 DDKID-LFKNTNDFAIKKLIND-----DTKDMGLKLLSTGLVQNFNTI 1574

RESULT 8
MSPI_PLAPP STANDARD; PRT; 1726 AA.
AC P50495;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (GPI95).
GN MSP-1.
```

```
OS Plasmodium falciparum (isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270,
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Siddiqui W.A.,
RT "Plasmodium falciparum: gene structure and hydropathy profile of the
RT major merozoite surface antigen (Gp195) of the Uganda-Palo Alto
RT isolate.";
RL Exp. Parasitol. 67:1-11(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M37213; AAA29611.1; -
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF0008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5859CEBF9A026 CRC64;

Query Match 3.6%; Score 312.5; DB 1; Length 1726;
Best Local Similarity 19.9%; Pred. No. 9.6e-05;
Matches 350; Conservative 286; Mismatches 600; Indels 523; Gaps 90;

Qy 51 DINNLDNLYNTYKSGRNKALKKFKKYLTWVLELKNLSLTPVEKNLHFTWIGQINDT 110
Db 154 DLKHRVRYLFT-----IKELYPELFDLTNMLTLCD-NIH----- 189
Qy 111 AINY-INQWQVNSDYTVKPYVDSNAFLNTLTKTIVESATNTLE-SFRENLDNPEFDY 168
Db 190 GFKYLIDGYEIN-----ELLYKLN-FYFDLRLAKLNDVCAVNDYCOIPFNKIRANELD 243
Qy 169 NK----FYRKMEIYDKQKHFIQYKSOIEENPEFIIDNIKTVLSNEYSKDLKALNKY 224
Db 244 LKLVFGYRKPLDNKDNVGMEDYIKN-----KTTIAN-----INEL 282
Qy 225 IEESLNKITANGNDIRNLEK---FADEDLVRLYNQELVERWNLAASDILRISMLKBDG 281
Db 283 IEGSKKTIQDNKADNBEKKLYQAQYDL-SYNKQLEEAHNLISVLE-KRIDTLKNE 340
Qy 282 GVLVDVILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAINMKYKYEIPGYTSKPFMDLDE 340
Db 341 NIKELDKINEIKNPPPPANSNGTNTLDDKNKKIEEHEE--KIKE-TAKTIKFNIDSFT 397
Qy 341 EVQSFESALSSKSKDGBIFLPLDDIKVSPLEVKIAFANNVINQALISLSDSYCSDLVI 400
```

Db	398	D-PLEBYLREKNKVDVTPKSDPYKVOIQKVPNGIVPLPLUTDTHNSLAAD---	453
Qy	401	NOIKNRY-KILNDNLNPSINEGDTFNTMKIFSDKLASINEDNMFMKITYLNKVGFA	459
Db	454	NDKNSYGDLMNPDTEKINE-----KIITD-----NKRKIFI-----NNIK----	490
Qy	460	PDVSTINLSPGVYTGAYQDLMFKDKNSTNIHLLPELANFPFPKTKISQLTEQBITSL	519
Db	491	-----KQIDLEEKINHTEQNKCLL	511
Qy	520	WSFNQARAKSOPREYKGYEGALGEDDNLDPQNTVLDKDYVSKYKTLSSMKTRNKE---	576
Db	512	EDYE--KSKDYBELLEKFEY-----MKFNNN--FDKDDVD--KIFSARYTINVEKOR	558
Qy	577	YIH-----YVQJQGDOKISYEASCNLFSDPYSSILYQKNIEGETAYTYVV--A	624
Db	559	YNNKFSSNNSVYNQKLKALS-----LEDYS--LRKIGSEKDNHYLLTKTGL	607
Qy	625	DAEIKEIDKYRIPYQISKNK--IKLTFIG--HGKSEFNTDTPANLDV-DLSSEIETIL	679
Db	608	EADIKKLTE-----EIKSSKNILEKNFKGLTHSAN-----ASLEVDIVLQVQKVL	655
Qy	680	NLAKADISPKYIELN-----LGCNMFPSYSIYAEETPGK-----LLLKIKDRVSELP--	728
Db	656	LIRKIEDLRK-IELFLKNAQKDSIHVPNIYKPNKPEPYLLVLKVEDKGLKEFPKVK	714
Qy	729	-----SISQDSITVSANQYE-----VRINEEGKEIILDHSGKINKEISIKOI	772
Db	715	DMLKQEAQLSSITQPLVAASETTEDGGHSTHLSQSGETEVTEETEE---TEETV--GH	769
Qy	773	SSKEYISFNPKENKIIVSKYLHELSTLQEIIRNNANSDDILEKKVMLTECEINVASN-	831
Db	770	TTTTITLPPKEKVVENS-----IEHKSNDNSQALTKTIVLKKLDEFLTQSY	817
Qy	832	-----IDRQIVEGRIEAKNLTSDSINYIK-----NEFKLIESIDSL-YD	871
Db	818	ICHKYILVSNSSMDQKLEVLNLTPEBENELKSCDPLDLFNTONNIPAMYSYDSSMND	877
Qy	872	LKHQGLDDSHFISPEDISIKTENGFRIRFNKGTGNSIFITEKEI-----	917
Db	878	LQH-----LFFELYKEMYYLHLKKEENHIKKLEBQKQITGTSSTSSPGNTTV	927
Qy	918	-FSEYATHIS--KEISNIKDTIFDN-----VNGKLVKVNLDAAEHVNTLNSAFFIOSLIE	970
Db	928	NTAQSAHNSONQOSNASSTNTQNGVAVSSGPAVVEESHDLPTVLVLSI\$NDLKGIVSLLN	987
Qy	971	Y-NTTK-ESLSNL\$VAMKVQVYAQLFSTGLNTITD-----ASKVVE--LVSTALDETI	1019
Db	988	LGNKTKVPNPLTISTTEMBKFYENILKNNDTYFNDDTIKQFVKSNSKVIITGLTQKNALN	1047
Qy	1020	DLPLTLEGLPIIATIIDGVSL-----GAAIKELSETNDPLLROETEAKI	1064
Db	1048	DEIKKLKDTQLSPDLNKNYKLLKDLRFLNKKKELGQDKMQKKUT-----LLKEQLESKL	1102
Qy	1065	GIMAVNLTAASTAIVT\$ALGIAGFSILL--VPLAGISAGIPSLVNNNELLODKATKVID	1122
Db	1103	NSL-----NNPHNVQLQNSVFFPNKKKEAEIAETENTLENTKILLKHKGVLKY	1150
Qy	1123	YFKHISLAETGAFTLLDDKIIMPQDOL-----VLSEIDFN--NNSITLKGCEIWRABG	1174
Db	1151	YNGESSPLKT-----LSEYSIQTEDNYANLEKFRVL\$KIDGKLNLDNLHKGKCL-----	1199
Qy	1175	GSGHTLTDDIDHFPSSSITVRK\$PWL\$SYDVNLNIKKEIDF\$KDL\$VLPNAPNRFVGYEM	1234
Db	1200	-----SPLSS-----GLHQILTELKEVI--KNKNVTGNSP-----	1227
Qy	1235	GWTPGFRSLNDGCKLLDRIDRHYEQFYWRYFAFIADALITKL-----KPRYEDTNNVRIN	1290
Db	1228	-----SENNKVVNEALKSVEN-----FLPEAKVTIVTPQPDVTP\$PLSVR	1269
Qy	1291	LDGNTRSFIVPVITTEQIRKNLSY\$FSGG\$Y\$LSL\$SPYNNMIDMLV-----END	1342

Db 1270 VSGSGS-----TKETQJPTSGSL-TELQVVLQNYDEDDSLVLPFGESDND 1320

Qy 1343 TW- - - - -IDVNVVKNITIEDEIQ- - - - -KGELFENTILS-KLNIEDN 1380

Db 1323 EYLQDVVTGEAISVTMDNILSGFENEYDVLYPLAGVYRSLKKQIEKNIFTFNLLND- 1381

Qy 1381 KIILNNH--TINFVGDINEN-----RPSITLTFSEILEDINIHHIDLVSYSKYLLSGNCM 1434

Db 1382 --ILNSLRKKRKYPLDVLSDLMQPKHISSEYIIE-----SFKLLNSEQKN 1427

Qy 1435 KLINSSDIQOKIDH-IGFNGEHQY-- - - -IFYSVIDNETKNGFI- - - - - 1475

Db 1428 TLLASYKVIKESVENDIKFAQEGISYEVKVLAKYKDDLESIKKVIKEEKKFPSSPPTP 1487

Qy 1476 -----DYSKKEGLTFAFSGESIIRNIYMPDSNNLFYSSKDKOIRIINKGD- - - - - 1523

Db 1488 PSPAKTDEQKESKFLPLTN- - - - -IETLY- - - - -NNL- - - - -VKIDDDVLLNL 1527

Qy 1524 -VKLLIGNYFKDDMKVLSFTIEDTNTI--KLNGVYLDENGVAQILKFMNKAALNTSN 1580

Db 1528 KAKINDCNVKEDEAHVKIT-KLSDLKAIDDKID-LFKNHNDFAIKKLIND-----DTKK 1580

Qy 1581 SLNNFLESINIKNIFYNNL 1599

Db 1581 DMLGKLLSTGLVQNFPNTI 1599

RESULT 9

MSPL1 PLAPFW

ID MSPL1 PLAPFW STANDARD; PRT; 1639 AA.

AC P04933;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMMSA) (P195).

GN MSP-1.

OS Plasmodium falciparum (isolate Wellcome).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5848;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86014355; PubMed=2995820;

RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,

RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,

RA Freeman R.C.

RT "Primary structure of the precursor to the three major surface

RT antigens of Plasmodium falciparum merozoites.";

RL Nature 317:270-273(1985).

RV [2]

RN REVISIONS.

RA Holder A.A.

RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor

CC (PTM).

CC -!- POTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42

CC kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X02919; CAA26676.1; --

DR PIR; A24594; A24594.

DR PIR; S05603; S05603.

DR PDB; ICEJ; 28-MAY-99.

DR InterPro; IPR006209; EGF\_like.

DR Pfam; PF00008; EGF; 1.



KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;	
FT	Transmembrane; GPI-anchor; 3D-structure.	
FT	SIGNAL	1 19
FT	CHAIN	20 1639
FT	CARBOHYD	116 116
FT	CARBOHYD	268 268
FT	CARBOHYD	764 764
FT	CARBOHYD	768 768
FT	CARBOHYD	783 783
FT	CARBOHYD	844 844
FT	CARBOHYD	920 920
FT	CARBOHYD	964 964
FT	CARBOHYD	1058 1058
FT	CARBOHYD	1165 1165
FT	CARBOHYD	1174 1174
FT	CARBOHYD	1445 1445
FT	CARBOHYD	1526 1526
SQ	SEQUENCE	1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;
Query Match		
Best Local Similarity 19.5%; Pred. No. 0.00012; Length 1639;		
Matches 351; Conservative 280; Mismatches 554; Indels 618; Gaps 94;		
QY	36 NMSESVVEKYLKLDINNLDTNLTNYKKSGRNKALKKFKREYLTMEVLEKNNSLTPVE	95
DB	112 NPSDNSSDSDAKSYADLKHVRNYLLT-----IKELKYPQLFDLTNNMLTICD	159
QY	96 KNLHFIWIGQINDTAIFY-INQKDVNSDYTVKFPYDSNAFLNTLTKTIVESATNTYL	154
DB	160 -NIH-----GFKYLIDGYEIN-----ELLYKLN-EYFOLLRAKLDVANCYD	201
QY	155 E-SFRENLDPPFDYNNK-----FYRKMEIYDQKHFDIYKSOEENPEFIIDNIITY	209
DB	202 QIPFNLIKIRANELDVLKLVFGYRPLDNKONVGKMDIYKNN-----	245
QY	210 LSNEYSKOLEALNKYTESLNKITANNGNDIRNLEK-----FADEDLVRLYNOELVERNWL	265
DB	246 ----KKTIIENIELIEES-KKTIDCKNKATKEEKKLYQAQYDL-SYNNQLEEAHL	298
QY	266 AASDLIRLSMLKEDGGVLDVDILPGIO-PDLFSKINPKDSITNTSMEMIKLEAIMKYK	324
DB	299 ISVLE-KRIDTLKKNENIKELDKINEIKNPPANSNGTNTL-----LDNKKKIE	348
QY	325 EYIPGYTSKNFMDDEVQVSFESALSSKSDKSEIFLPDDDKVSPLEVK--IAPANNVS	382
DB	349 EH-----EKEL-----KEIAKTIKFNIDSFTDPLELEYLYREKNNKI	386
QY	383 INQALISLSDYSYCSDLVINOIKN--RYKILNDLNPSINEGTDFTNTWKIFSCLKASIN	440
DB	387 DISAKVETKES---TEPNEYGVVTPPLSYNDINNALNELNSFG-----	427
QY	441 EDNMFMKIKITVLKVGFPADVRSTINLSGPGVYTGAYQDLMFKDNSTNIHLLEPELRN	500
DB	428 -----DLINPPDYT-----REPSKNIY-TDNERKK	451
QY	501 FEPFPTKISQLTEQ-EITSLMSFNQARAKSQFEYKKGVEGALGEDDNLDPQNTVLDK	559
DB	452 F-----INEIKKIKI-----EKKKIESDKSYEDRSKSLND-----ITK	486
QY	560 DYVSKKILSSM---KTRNKEYIHYIYVQLQGDKISYEASCNLFSKDPYSILYQK-NIEGS	615
DB	487 EY-EKLLNEIYDSKFNNNDILTNPEKMGKRYKVE-KLTHHTNTFASYENSKENLE--	541
QY	616 ETAYYYYVADAIEIKDKVRIYQISNKENIKLTFIGHCKSEFNTDTFANLDVDSLSSEI	675
DB	542 -----KLTALKKYMEDYSL-----RNIVV-----EKELKYKKNL-ISIKENBI	578
QY	676 ETILNLAKADISPKYTEINLGCNMFPSYIYAEETYPGKLLKIKORVSELMPSISQDSI	735
DB	579 ETILVENIKDEE-----QLFEKKITDKENKPKDEKILEVSDIV-----	615
QY	736 TVSANQYVRINEBEGKEILDHSGKWINKESIIKDISSKEYI---SFNPKENK-----I	787
DB		

DB	616 -----KVQV-----QKVLLMKNKIDELKTKTQILIKNLVELKHINHPNSYKQENKQEPYVL	664
QY	788 IVSKYLHELSTLLOEIRNNANSSDIDLEKKVMLTECEINVASNIDRQIVEGRI-----	841
DB	665 IVLKKEIDKGVFMPKVESLINE-----EKNIIKTEGQ-----SONSEPSFTEGITGOATT	715
QY	842 ---EAKNLTSDSINTVKNFEFLIE-SISDSLYDLKHQ-----NGLDSDHFI SPED	888
DB	716 KPGQAGSALGSDSVQAQAEQKQAPVPVPVPEAKAQVTPPPAPVN-----	763
QY	889 ISKTENGFRIRFINK--ETGNSIIFETEKEIFSYATHISKIENIKOTIPDNVNGKLVK	946
DB	764 -NKTENSKLDYLEKLYEFLNTSYI-CHKYILVSHST-----MNEKILK	805
QY	947 KVNLDAAHEVNTLSAFPIQSLIENYTTKESLNSLVAMKVQVYAQLFSTGLNTITDASK	1006
DB	806 -----QYKITKEESKLSGC-----	820
QY	1007 VVELVSTALDETIDLLPTLSEGLPIIATIDGV--SLGAAIKELSE-----TND	1053
DB	821 -----DPLDLLFNIGNNIPMYSMFDSLNLSQLFMEIYKEMVCNLYKLKOND	870
QY	1054 PLLRQETAEKIGIMAVNLTAATAITVSALGIAGSFIILLVPLAGISAGIPSL-VNNELI	1112
DB	871 KIKNLLBEAK-----KVSTSVKTLIS-----SSSMQPLSLTPQDKPEVSANDDT	914
QY	1113 LODKATKVIDYFKHI-SLAETEGAF-TLLDDKIIMPQDDLVLSEID-PNNNSITLGRCEI	1169
DB	915 HSTNLSNLSKUFENILSLGKNKNYQELIGOKSENFEYKILKOSDTPYNESFT-----	970
QY	1170 WRABGGSGHTLTDIDHFFSPSITYRKPMLSI-YDVLNIIKKEKID--FS-----	1216
DB	971 VKSRADDINSINDSKRKKLEEDINKLKTQLSFDLYNKYKLLERLFDKKTGVKYKM	1030
QY	1217 -KOLMVLN-----NAPNRVF-----GYEWGTFGRSLNDGDKLDR	1253
DB	1031 QIKKLLTKLEQLESKLSLNNPKHVLFNSVFFNKKKEAEIAETE--NTLEN--TKILLK	1086
QY	1254 IRDHVEG-----QFYWRVFAFTADALITKLKRYEDTNNVRINDG	1293
DB	1087 ---HYKGLVKYNGESSPLKTLSESIQTEDNYASLENFKVLSKLEGLKD--NLNLEK	1140
QY	1294 NTRSFIVP-----VITTEQIRKNLSYFSGSGSYLSLSPYNNIDNLN-----	1338
DB	1141 KKLVSLSGLHLHLLAEKLVKKNYKNT-----GNSPSENNTDVNNALESYKFLP	1190
QY	1339 -----VENDTWI--DVNVVKNITIT-----ESD	1360
DB	1191 EGTDVATVWSRSGSDTLLEQSQPKKPASTHVGAESNTITTSQNVDDEVDVIVPIFGESE	1250
QY	1361 E-----IQGE-----LIENILSKLNIE-----DNKILNHNHTINPYGD	1394
DB	1251 EYDDELGVVTEAVTSPVIDNLSKIENEYEVLYLPLAGYRSLKQLENNVMTFVN	1310
QY	1395 INESNRPFISLFILEDINIIIEIDLVSYSKILLSGNCM-----KLIENS	1440
DB	1311 VKD---ILNSFKNRKNFKNVLESGLI--PYKOLTSNNYVVKPYKFLNKKROKFLSSY	1365
QY	1441 SDIOQKID-HIGFNGEHQKIYFYVIDNETKNGFIDYSKKEGLFTAFBSNESIIRNTYM	1499
DB	1366 NYIKSDIDTINFANDVLGY--YKILSE--RYKSDLD-SIKKYINDKQGENEK-----	1415
QY	1500 PDSNNL-FIYSS-KDLKDIRIINKGDVKLLIGNYFKDDMKVSLIS-----FTIEDNTIKL	1552
DB	1416 PFLANNIETLYKTVNDKIDLFVHL-EAKVLNNTYKESNVEVKIKELNVLKTIQD-----	1470
QY	1553 NGVVLDEN--GVAQILKFMNNAKSALNTSNLSMNFLESINI-----KNIFYNNLDNPIEF	1605
DB	1471 ADFKKNNFVGIADL-----STDYHNHLLTKFLSTGMVFENLAKTVLNSLLDGNLQ	1523
QY	1606 ILD 1608	
DB	1524 MLN 1526	



```

RESULT 10
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT: 2136 AA.
AC P09375;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
OC Marchantiaceae; Marchantia.
OC NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohshima K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.
RT "Chloroplast gene organization deduced from complete sequence of
RT liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574 (1986).
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; X0465; CAA28078.1; -
DR PIR; S01591; A05037.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR Pfam; PF00004; AAA_1.
DR SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CF61197 CRC64;

Query Match
Best Local Similarity 3.5%; Score 306.5; DB 1; Length 2136;
Matches 387; Conservative 271; Mismatches 629; Indels 623; Gaps 102;

QY 27 ILNALLEEYHNMSESSVEKYKYLKDINNLT-DNYL-----NTYKSGRKNALKKPKLEYLT 80
DB 309 IPKNLQNP-NESDKKLIIESFFLLKIKGNLYFKNYIEFVTWQSYKKDCLD--FNKNELNN 365

QY 81 MEV-LELKONSUTPVEKNLHFTWIGQINDTAI-----NVINQWQVNS--DYTVKPVYDS 133
DB 366 SEIYKIEBELFSDIYKFSKLYLEGKSKTKTIKQSFNNIYKKLNSFNFTIYFDS 425

QY 134 NAFLINTLTKKTIVESTATNT--LESFRENLDPEFY-----NKFYKRMELIYDKKHFI 187
DB 426 NLLFDWLKKNY--INKPLKGF-----LIYSSINQF-----ILFFKQKSK 468

QY 188 DYYKQIEENPEFIIDNIITKYLVS--NEYSKDL-----EALNKYIEESLNKIT 233
DB 469 SFNKNLVKQNSKDVITNVPNSKENKEINNFPSKIYAFPEILSINEIDNKFV---INKIS 525

QY 234 ANGNNDIRN-----LKPFADEDLVRLYNQELVERWN-----LAAASDILRIS-ML 277
DB 526 LKNINKKQKRYFLNKRKSSDNFRFINLWKIKNYSQVFNNSFLNPAFELQOYYL 585

QY 278 KEDGGYLVLDVILGQPLDPLKSIKPKDS-----ITNTSEMIKLEAIMK----- 322
DB 586 KGNILP-----FKKLEVFNSFFYQYKCKKLNIFLKPFASLEKILKGRNK 632

QY 323 -----YKEYIPGVTSGNFDMLDEBQVRSFESALSSKSKSEIFLPLDDIKV----- 368

```

```

DB 633 KFTISIKLKFYKKNLNGEYKTESQILQNEKELNKKRKNQFNP--NIKILSFYNS 690
QY 369 SLEVKIA---PANNVINQALISLKDSYCSDLVINQIKRKYKILNDNLNPSINBGTDFN 425
DB 691 SKGNIYQKQYFFNKNLNNKLITWK-----KISNKLIVSNSEYKNKIWN---K 736
QY 426 TTMKIFSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGPGVVYTGAYQDL---- 481
DB 737 KMKFFSFSKNSVL--DTFFFNKKSFNITVIF--DKLKKIQLN-----FQEQIKL 784
QY 482 ----LMPKDNSTNIHLSPELRNFPFKTKI---SOLTEQETSLSWFSNQARAKSQFBEY 534
DB 785 NCFSLFF--NSKNI-----KKTIPKNSYFINENLTTPSFND----KEFNIF 826
QY 535 KKGYPEGALGEDNDLDAQNTVLDKDYVYSKTLSS-----MKTRNKVHYHYIV 582
DB 827 FLELFISBINNDFLMRFPFKYLYRYKDKBELFNPENRQLQLQNFPEKTKILTIDF-- 884
QY 583 QLQGGKISYEASCNLFSDPYSSILYQKNISETAYY-----YYVADA-----EIKE- 630
DB 885 -LQDEPANY-----NNRFIFLHEKTKIKNNLLYLRLKIFLKDKNFLLINEIKSF 935
QY 631 IDKYR---IPQISNKRNIKLTFFIGHGSEFNTDTFANLVDVLSSEIETILNLAKADIS 687
DB 936 IEKKNLFIKSQSLNVLVKSNSY-----KPPDNIFNFHFLKQKEKNIEILN-----N 983
QY 688 KYIIEINLL-----GCNMFYSIYABET-----YPGKL 715
DB 984 QNYFEKSLKTKYLNKLNLSNSKFSYKIFIFQLNLNLKNKYKTFQWISLIFYSKNL 1043
QY 716 LKIKDRVSELMPISQSDISITVSANOYEV--INEGKREILDHSGKWINK----- 764
DB 1044 NYKIQNKIEKNYCNKN---ISYKKKIKITVPEKKNLFTQNNSWPFFLEWEVNTYI 1100
QY 765 -----EESIIDISSKISYISPNKKNKIIVSKYLHELSTLQRIINNANSDDILEKKV 819
DB 1101 LIQIIQETFFQITDVLVFP-----KKKIIIEKN-----LKFFLK-----SKKI 1138
QY 820 MLTEC-----EINVASNIDRQIVEGRIEAEKNLTSDSINYIKNEPKLIES 864
DB 1139 SLKTLSTFHNFKLKNLRFNFEINKNKY-----LNLFLMSDFNLINN 1180
QY 865 TSDSLY-----DLKHQN-----GLDSDH--FISFEDISKTEGNGPRIPINKET-G 906
DB 1181 -CNNLYWIFSLVIFLYQKIFSIILIGSDCFHLWKVPEILQYLTDRSRSIYFKLTR 1239
QY 907 NSFIETEKEIPSEYATHISKEISNIKOTIDPNVNGK--LVKKVNLDAHAHEVNTLNSAFP 964
DB 1240 NKTALNKTENLLSYFQNLTHYITNFKVLLTKKLNKKNLNNKTLDLSRRKRKL---L 1295
QY 965 TOSLI-----EYNTTKE-----SLSNLSVAMKVQVYAQLFSTGL--NTITDA 1004
DB 1296 VQSLTITHNKIQNYGFELNSKQFTSYFYQYQITNOQGLLYFOYLAQFPQKNLINSLDIA 1355
QY 1005 SKVBE-----LVSTALDET--IDL-----LP-----TLSEGLPIATIDGVSGLAA 1044
DB 1356 NKWIVFSFWHKIFSSQKLRQTNWIELGQNIYVPVLPQGLSYSGKILLIGPIETGYSY--L 1413
QY 1045 IKEL--SETNDPLRQIEIAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGI 1103
DB 1414 IKNLAAESVPLFKISIN-----KLLYNKPDVITE-----SWNNILIESLRL-- 1456
QY 1104 PSLVNNELLQDKATKVIDYFKHI-----SLAETEGATFLDDKLIIMPQDLDVLS- 1153
DB 1457 ----NLTDFAKMSPCIWIQNIHQNLNRLNTQVESDPTFL-----GILLKY 1502
QY 1154 -EIDFN-----NNSITLKGCEIWAEGSGSHTLTDIDHFFSSP-----SITVR 1196
DB 1503 FOTDFSKTKKNIIIV-----GSTH-LPKKVDPALISFNRLDKIINVLRFNISR 1551
QY 1197 KPWLSIYDVLTAKKEKIDFSDKMLVLPNAPNRVFGYEM-----GWTGFRSL 1243

```



```
QY 751 KREILDHGKWKINKESIIKOISKEYI---SFNPKENK-----IIVKSKYLHELSTLLQ 802
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 611 QVLLANKIDELKTKTQILKVKLKHNIHVPSYKQENKQEPYVILVKKEIDKLKVFNP 670
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 803 EIRNVANSSDIDLEKVMLTCEINVASNIDQIVEGRI-----EAKNLTSDSI 852
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 671 KYESLINE-----EKNINTEQO-----SDNSPSTEGEITGOATTKPGQAGSALGDSV 721
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 853 NYIKNEFKLIE--SISDSLVDLKHQ-----NGLDDSHFISFEDISKTENGFRIRFNK 903
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 722 QAAQEQKQAQPPVPVPEAKAQPPTPPAPV-----NKTENVSKLDYLBK 768
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 904 --ETGNSPIETEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKKNVLDAAHEVNTLNS 961
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 769 LYEFLLNTSYI--CHKYILVSHST-----MNEKILK----- 796
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 962 AFFIQSLTEYNTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETIDL 1021
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 797 -----QYKITKEESKLSK-----DPLDL 816
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1022 LPTLSEGLPIIATIDGV--SLGAAIKELSE-----TNDPLLRLQETEAKIGIMA 1068
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 817 LENIQNNIPVMYSFMSFDSLNSLSOLFMEIYKEMVCNLYKDKNDKIKILLSEAK----- 871
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1069 VNLTAASTAIVTSALGIASGFSILLVPLAGISAGIPSL-VNNELILQDKATKVIDYFKHI 1127
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 872 ----KVSTSVKTL-----SSSQPLSLTPQDKPEVSANDTSHSTLNNSLKLPENI 920
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1128 SLAETEGAF-TLLDDKIIMPODDLVLSEID-FNNNSITLKGCEIWAEGSGHLYTDDI 1184
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 921 LSLGKNQNIQBLIGOKSSENFYKILKDSDTFYNESEFT----NFVKSADDSINSLNDES 976
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1185 DHFFSSPSTTYRKPLWSI-YDVNLNIKKEKID--FS-----KDLMLVP----- 1223
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 977 KRKLEEDINKUKTLQLSFDLYNKYKLERLFDKKTGVKYKQWKLTLLKLSQLESK 1036
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1224 ----NAPNRVF-----GYEMGWTPGFRSLDNDGTLLDRIRDHVEG----- 1260
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1037 LNSLNNPKHVLPNSVFFNKKKEAEIETE--NTLEN--TKILLK--HYKGLVKYNGE 1089
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1261 -----QTVWRYPAFTADALITLKKPRYEDTNVRINLGNWTRSPFIVP-----VI 1303
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1090 SSPLKTLSESIQTEDNYVASLENFKVLSKLEGLKD--NLNLEKKKLSYSSGLHLIA 1146
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1304 TTEQIRKNLSYFYSGSGSYSLSPYNNIDNL----- 1338
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1147 ELKEVIKKNKYT-----GNSPSENNTDVNNALSESYYKFLPEGTDVATVSESQSD 1196
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1339 -----VENDTWVI--DVDNVVKNITI-----ESDE-----IQKE-- 1366
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1197 TLEQSQPKKPASTHVGAESNTITTSQNVDDVDVVIPIFGESEEDYDLGQVVTGEAV 1256
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1367 ---LENILSKUNIE-----DNKIILNNHTINFYGDINESNRFSITFSL 1409
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1257 TPSVIDNLISKIENEYVLYLKPLAGVRSYKQLENNYMTFNVVKD---ILNSRFNKR 1313
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1410 EDINIIIRIDLVSYSKILLSGNOM-----KLIENSDDLOQKID-HIGENG 1454
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1314 ENFKNVLSDDL--PYKDTSSNYVVKDPYKFLNKEKRDKFUSSYNYIKDSITDINPAN 1371
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1455 EHQQYIFYSYIDNETKYNGFIDYSKKEGLFTAEFSGNESIIRNIYMPDSNNL-FIYSS-KD 1512
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1372 DVLGY--YKILSE--KYKSDLD-SIKKYINDKQGENEK-----YLPFLNNIETLYKTVND 1421
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1513 LKDRIIRINKGVKLLIGNYFKDDMKVSL-----FTIEDTNTIKLNGVYLDEN--GVAQI 1565
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1422 KIDLFPVHL--EAKVLNITYEKSNEVKIKELNYLTKTIQD---KLADFKNNNFVGIADL 1476
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1566 LKFMNNAKALNTSLNMFLESINI-----KNIFYNLLDPNIEFILD 1608
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1477 -----STDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQGMNL 1517
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 12
SPOF SCHPO
ID _SPOF SCHPO STANDARD; PRT; 1957 AA.
AC Q10411; QOUSES;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
GN SP015 OR SPAC1F3 06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxID=4896;
RX [1]
SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20107136; PubMed=10639340;
RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Sp015p is localized
RT to the spindle pole body and essential for its modification.";
RL J. Cell Sci. 113:545-554(2000).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy I., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch B.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas R., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benico J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RN Nature 415:871-880(2002).
[3]
SEQUENCE OF 705-871 FROM N.A.
RP STRAIN=968 h90;
RC MEDLINE=2023868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: Has a role in the initiation of spore membrane
CC formation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Spindle pole body.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```





OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 28383 / FL100;  
 RX MEDLINE=92079907; PubMed=1745235;  
 RA Kormanec J., Schaaff-Gerstenschlaeger I., Zimmermann F.K.,  
 RA Ferecko D., Kuentzel H.;  
 RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the  
 RT highly repetitive 313 kDa NMI protein.";  
 RL Mol. Gen. Genet. 230:277-287(1991).  
 CC -!- FUNCTION: CONTROLS NUCLEAR MIGRATION. NMI SPECIFICALLY CONTROLS  
 CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-  
 CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC  
 CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR  
 CC ENVELOPE.  
 CC -!- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT  
 CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH  
 CC FLANKING DOMAINS OF THE TANDEM REPEATS.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X61236; CAA43554.1; --  
 DR SGD; S0002557; NMI1.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00169; PH; 1.  
 DR SMART; SM00233; PH; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 KW Repeat.  
 FT DOMAIN 593 1384  
 FT REPEAT 593 656 1.  
 FT REPEAT 657 727 2.  
 FT REPEAT 728 798 3.  
 FT REPEAT 799 862 4.  
 FT REPEAT 863 926 5.  
 FT REPEAT 927 990 6.  
 FT REPEAT 991 1054 7.  
 FT REPEAT 1055 1118 8.  
 FT REPEAT 1119 1182 9.  
 FT REPEAT 1183 1246 10.  
 FT REPEAT 1247 1310 11.  
 FT REPEAT 1311 1374 12. (INCOMPLETE).  
 FT REPEAT 1375 1384 13  
 FT DOMAIN 2573 2683 PH.  
 SQ SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;  
 Query Match 3.4%; Score 292; DB 1; Length 2748;  
 Best Local Similarity 19.8%; Pred. No. 0.001;  
 Matches 416; Conservative 346; Mismatches 686; Indels 652; Gaps 117;  
 QY 5 NKAQLQGVYKPRIOEDVVAIINALBEEVYHNMSESSVVEKYLKLDIN-NLTDNYLNTY 63  
 DB 118 NEAKNQIKSLK-QIKE-----SLSDKIEELTN-QKSPMKELDSTKLANWDLKSLTNLS 171  
 QY 64 KSGRNVKALK-----FKVLTMEVLELNKNSLTPVEKNLHFTWIGQNDFAINVI 115  
 DB 172 MECRLKELKKTKESWNEDESKLLKTDLEILT-----KN-----GMENDLSQKL 221  
 QY 116 NQWKDYN--SDYTVKFDVSNAPLINTLKKTI-VESATNTLE--SFRENLDPEFDYK 170  
 DB 222 HYDKEISELKERLIDLNENDRLILSVSLTSEINLSQNRTERIKIQLODOKASIS 281  
 QY 171 FYKRMEIIVDKQKFDYVYKSIQENPE-----FIIDNIKIYLS-NEYS 215

Db 282 LKRRVQKYYQKQ-HTSD---TTVTSDDPSDEGTTSEEDIFDIVIEIDHMIETGSPSVEDIS 337  
 QY 216 KOLEALNKYIEESLNKITANNNDIRNL-----EKFADEDLVRLYNQELVERWNLA 268  
 DB 338 EDL--VKYSEKN-NMILLSN-DSYKQLLQKSEGASKPKODELM--TKVAENLNMTAL 390  
 QY 269 SDILRISMLK--EDGGVYLDVILPGIOPDLFKSINKPDSITNTSWEMI--KLEAIMKY 323  
 DB 391 PNDNYSKFEFSLESHIKYLEASGYKULPLEEFENLN--ESLSNPSYNYLKEKQALKK- 447  
 QY 324 KEYIPGYTSKNFMDLDEEVQSFESALSSK-----SDKSEIFLPDDDIKVP--L 371  
 DB 448 --IP-IDQSTFNLL-KEPTIDFLPLTSKIDCLIIPTKDYNDLFESVKVPSIEQMKKCL 502  
 QY 372 EVKTAFAWN-----SVINQALISLKDSYCSYSDLV--INQIKRKYI 409  
 DB 503 EAKNDLQSNICKWLEERNGCKWLSNDLYFSMVNKIETPSK-QYLSDKAKEDQVLDIDTKA 561  
 QY 410 LNDNLNPSINEGTDFTNTMKIFSDKIASISNED---NMFMFI---KITNLYKVGFPADV 462  
 DB 562 LEGLKNPTI---DFLREKASADYLL-LKEDYVSPSELYLVEHAKATNHHLLS----- 611  
 QY 463 RSTINLSGPGVYTGAYQDLMPKONSTNIHLEPELRFEPFKTKISQLTQEITSLWSF 522  
 DB 612 -----DSAYEDLVKCKEN-----PDM--EFLKEKSAKLIGHTVVS----- 643  
 QY 523 NOARAKSOFEBYKGYKYPGALGEDDNLDF-----AQNTVLDKQVSVKLLSSMKTRNKE 576  
 DB 644 NEA-----YSELEKKL-----EQPSLEYLVEHAKATNHHLLSDSAYEDLVKCKENPOME 692  
 QY 577 YI-HYIVOLQGDKISYEASCNLFSDKDPYSSILYQKNIQEGSETAY-----YYVADA 626  
 DB 693 FLKEKSAGLHTVTVSNEAYSELQK--YSEL--EKEVEQPSLAYLVEHAKATDHHLLSDS 748  
 QY 627 EIKEIDKYRIPYQISNKENIKLTFIGH-----GKSEFNTDTFANLD--VDSLSEIET-- 677  
 DB 749 AYEDL-----VKCKENPDVEFLKEKSAGLHTVVSSEYSELQRYSELEKEVEQPS 800  
 QY 678 ---ILNLAKADISPKYIEINLLGCMFYSYIAYETYPCKLLKLLKIDRVSELMPSISODS 734  
 DB 801 LAYLVEHAKA-----TDHLLSDSAYEE-----LVKCKENPD--MEFLKEKS 840  
 QY 735 I-----TVSAN---QYEVRIINEEGKREILDHSGKWINKEESIIKDISSEYISFNPKNK 786  
 DB 841 AKLHTVTVSNEAYSELEKLEQPSLAYLVEHAKA---TDHLLSDSAYEDLVKCK--KENS 895  
 QY 787 IIVKSKYLHELSTLQE--IRNANSNDIDLEKVM-----ELEKKLEQPSLAYLVEHAKATDHHLLSDSAY 949  
 DB 896 DV--EFLKEKSAGLHTVTVSNEAYS---ELEKKLEQPSLAYLVEHAKATDHHLLSDSAY 949  
 QY 821 --LTECEIN-----VASNIDRQIVEGRIEBAKNTLSDSINNYKNEFKLI 862  
 DB 950 EDLVKCKENPDMEFLKEKSAGLHTVTVSN-----EAYSELEKLEQPSLEYLVEHAKAT 1003  
 QY 863 ES--ISDSLXYDLKHQGLDDSHFISFEDISKTENGFRIRFINKET-----GNSIFTE 912  
 DB 1004 NHHLLSDSAY-----EDLVKCKENPDMEFLKEKSAGLHTVTVSNEAYSE 1047  
 QY 913 TEKEIFS---EYATHISKEISN--IKDTIF-----DNVNGKLVKYNLDAHEVNTLN 960  
 DB 1048 LEKKLEQPSLEYLVEHAKATNHHLLSDSAYBELVKCKENPDVEFLKEKSAGLHTVTV-VSN 1106  
 QY 961 SAFFTQSILIEYNTTKESLSNLSVAMKVQVYLAQLSTGLNTITDASKVVELYSTALDEITD 1020  
 DB 1107 EAY---SELEKKLEQPSLEYLVEHAK-----ATNHHLLSD-SAYBELVKCKENPDVE 1154  
 QY 1021 LLPTLSEGL-----PIIATIIDGVSL-----GAAIKEL---SET 1051  
 DB 1155 FLKEKSAGLHTVTVSNEAYSELEKLEQPSLAYLVEHAKATDHHLLSDSAYEDLVKCKEN 1214  
 QY 1052 NDPLLQRIEAKIGIMAVNLTAASPAIVTSALGIASGFSILLVPLAGISAGIPSL----- 1106  
 DB 1215 PDVEFLKEKSAGLHTVTVSNEA-----YSELEKKLEQ-----PSLAYLVE 1254



Qy	626	AEIKEDKYPVOISNKRNIKUTFIGHKSERNTDTFANLDV-----DSJSS-- 673
Db	451	AKVKEL-----NAKQKVLV--ECENDLQTLTKQRDLCLCRQIQYLLITNSVSNDS 497
Qy	674	-----EIEITILNAKADISPKYIEINLGCNMFYSIYABEYTGKLLKIKDRVSEL 726
Db	498	KGPIRKEEIOFINIMQEDDSTI-----TESDSQKVVTERRVLFPKNIJQOEKNAEL 549
Qy	727	MPISIQDSITVSANOYEVRIINEEGRILDHSGWINKEBSIIKDISSKEYISFNPKENK 786
Db	550	LKVVRN-----LADKLESK-EKKSQSLQKIESETVNEAKEAIIULKSEW-----DLESR 599
Qy	787	IIVKSKYLHELSTLLOBIRNANSSOIDLBEKKVMVLTECEINVASINDRQIVGRIEAKN 846
Db	600	IEELQKELEELKTSVP--NBDASYSNVTIKQ--LTETKRDLESQV--QDLQTRISQITR 652
Qy	847	LTSDSINYIKNEFKLIESIDSLDYDLKHQGLDDSHSISFE-----DISKTENG- 895
Db	653	ESTENMELLNKE---IQDLVDSKSDISIKLGKESKRILAEERFKLLSNLTLDITKAENDQ 709
Qy	896	FRIRFINKETGNSIFITEK--EIFSEYATHISK-----EISNIKDTIPDNVNGKLVKK 947
Db	710	LRKEFDYLO--NTILKQDSKTHETLNEYVCSKLSIVETELNLKE-----EQKLVRH 761
Qy	948	VNLDAHEVNTLNSAFFIQSLIEYNTTKESLNSLVAMK-----VOVYAGLFSTGLNTIT 1002
Db	762	LEKNLQKELNKLSP-----EKDSLURIMVTLQTLQKEREDELLESTRKSCQKKIDELE 813
Qy	1003	DASKVVELVSTALDETI-----DLLPTLSEGLPIIATIIDG----- 1038
Db	814	DALSELKETYSQKDDHHIKOLEEDNNSNIWQNKIEALKDYSESVITVSOKDQDIEKLQ 873
Qy	1039	-----VSLGAAAIKE-----LSET-NDPLLROEIEAKIGIMAVNLTAASTAI----- 1078
Db	874	YKVSLEKEIEEDKIRLHTYVMDETINDDSLRAKELEKS---KINLTDAYSQIKEYKDL 929
Qy	1079	--VTSALGIASGPSILIVLAGISAGTSPSLVNNELILODKATKVIDYFKHISLAETGAF 1136
Db	930	YETTSQ-----SLQOTNSKLDESFKDF---TNQIK 956
Qy	1137	TLDDKIIIMPDDLVLSIEDFN--NNSITLKGKEIWRAGEGSGHTLTDIDHFFSSPSITY 1195
Db	957	NLTDEKTSLEDKISLLKEQMFNLNNEIDLQK----- 987
Qy	1196	RKPWLSTYDVLNIHKKEDIFSCKMLVLPNAPNRVGYEMGWTGPFGRSLDN----- 1245
Db	988	-----KGMEKADPKFIRISILQNNNKEVAVKSEVESKLSKIQNDLDOQTIIYAN 1037
Qy	1246	-----DGTKLLDRIDH--YEGQFYWYPAFIADALITKLPKYED----- 1284
Db	1038	TAQNNYSELOQKHADVSKTISELREQLHTYKGVQ--KTLNLSRDQLENALIKENEKSWSSQ 1095
Qy	1285	-----TNVRI-NLDGNTSRFIVPVITTEQIRKNLSYFVGSGGSYLSLSPYNN 1333
Db	1096	KESLLEQLDLSNRIEDLSQNKLLYQIQIYTHADKEVNNSTNGPGJNILLTLRRED 1155
Qy	1334	I-----DLNLVENDTWIVDVNVKNITTESDEIQK--GELI-----ENILSKLN- 1376
Db	1156	ILDTKVTVVAERDAKMLRQKISLMDVELQDARTKLDNSRVEKENHSSIQQHDDIMEKLNQ 1215
Qy	1377	---IEDNKIILNNHTINPYGDINESNRFISLTFSEILEDINIIBIDLVSYSKILLSGNC 1433
Db	1216	LNLLRESNITLRNLEN-----NNNKKELOSELQKJQNVAPIESELTALKYSMQEKEQ 1271
Qy	1434	MKLI-----ENSSDIOQKIDHIGFNGEHOKYIFYSYIDN-----ETK 1470
Db	1272	LKLAKVEHVRHKKKSQDILEKHEQIS--SSDYEK--LESEIENLKEELENKERQCAEAEEK 1328
Qy	1471	YNGPIDYSKKEGLFTABFSNESIIRNI--YMPDSNNLFIYSKOLKDIRIINKGVDKLLIG 1529
Db	1329	FNR--LRROAOERLTKSLQDSLTFEOVNSLRDANKV--LENSLSBANARIIBELONAKVAOG 1386



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:31:53 ; Search time 40.625 Seconds  
(without alignments)  
7187.047 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1700  
Perfect score: 8677  
Sequence: 1 MNLVNAQLQRMVYKFRQ.....YLYGIDRYNVKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6733.5	77.6	2366	15	US-10-011-366-10
2	4458	51.4	2710	15	US-10-011-366-6
3	501.5	5.8	3169	15	US-10-114-170-257
4	310	3.6	2184	12	US-10-304-095-6
5	309	3.6	1639	15	US-10-087-464-10
6	288.5	3.3	1948	12	US-10-032-585-7611
7	286.5	3.3	1610	15	US-10-153-533-9
8	284	3.3	1881	12	US-10-032-585-7646
9	268.5	3.1	1847	10	US-09-801-368-56
10	266	3.1	1847	12	US-10-032-585-7538
11	257	3.0	6281	9	US-09-815-242-12996
12	256	3.0	2285	10	US-09-932-183A-2
13	253	2.9	1295	10	US-09-726-949A-1
14	250	2.9	4563	11	US-09-802-640-32
15	249	2.9	1706	9	US-09-864-761-46862

16	249	2.9	4563	10	US-09-870-759-128	Sequence 128, Appl
17	249	2.9	4563	12	US-09-751-708A-128	Sequence 128, Appl
18	247.5	2.9	3158	9	US-09-815-242-12611	Sequence 12611, A
19	247	2.8	1805	11	US-09-820-843A-73	Sequence 73, Appl
20	245.5	2.8	2437	9	US-09-815-242-5834	Sequence 5834, Ap
21	241	2.8	1786	10	US-09-742-096-3	Sequence 3, Appli
22	240	2.8	2871	15	US-10-146-473-41	Sequence 41, Appl
23	239.5	2.8	1169	12	US-10-241-596-20	Sequence 20, Appl
24	239.5	2.8	2025	9	US-09-815-242-5703	Sequence 5703, Ap
25	236.5	2.7	1979	15	US-10-205-823-419	Sequence 419, Appl
26	233.5	2.7	3878	12	US-10-080-608A-11	Sequence 11, Appl
27	233.5	2.7	3899	15	US-10-171-311-4	Sequence 4, Appli
28	233.5	2.7	3907	15	US-10-171-311-2	Sequence 2, Appli
29	233.5	2.7	3917	15	US-10-171-311-8	Sequence 8, Appli
30	233.5	2.7	3925	15	US-10-171-311-6	Sequence 6, Appli
31	232.5	2.7	2119	12	US-09-769-744A-28	Sequence 28, Appl
32	232	2.7	1598	12	US-10-032-585-7766	Sequence 7766, Ap
33	232	2.7	1781	10	US-09-995-749A-2	Sequence 2, Appli
34	229	2.6	1107	12	US-10-032-585-7483	Sequence 7483, Ap
35	228	2.6	2659	12	US-10-311-879-28	Sequence 28, Appl
36	227	2.6	1240	12	US-10-032-585-7366	Sequence 7366, Ap
37	227	2.6	1288	15	US-10-006-780-2	Sequence 2, Appli
38	224	2.6	1270	12	US-10-032-585-7127	Sequence 7127, Ap
39	220	2.5	3788	11	US-09-952-267-76	Sequence 76, Appl
40	219	2.5	1818	12	US-10-032-585-7616	Sequence 7616, Ap
41	218	2.5	3092	10	US-09-801-368-172	Sequence 172, Appl
42	217.5	2.5	1978	12	US-10-094-466-64	Sequence 64, Appl
43	217.5	2.5	2893	12	US-09-882-227-522	Sequence 522, Appl
44	217	2.5	2789	9	US-09-801-574-57	Sequence 57, Appl
45	216.5	2.5	1482	9	US-09-815-242-12484	Sequence 12484, A

ALIGNMENTS

RESULT 1

US-10-011-366-10  
; Sequence 10, Application US/10011366  
; Publication No. US20030054493A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; Kink, John A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/10/011,366  
; APPLICATION NUMBER: US/10/011,366  
; FILING DATE: 16-NO. US20030054493A1-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,310  
; FILING DATE: 23-OCT-1997  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:  
 NAME: Ingolia, Diane E.  
 REGISTRATION NUMBER: 40, 027  
 REFERENCE/DOCKET NUMBER: OPHD-01121  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2366 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
 US-10-011-366-10

Query Match 77.6%; Score 6733.5; DB 15; Length 2366;  
 Best Local Similarity 76.4%; Pred. No. 0;  
 Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

QY	1	MLNVKALQKQVYVVFRIQOEYVAIALNALEBYHNMSESSVVEKYLKLDINLTDNYL	60
DB	1	MSLVNRKQLEKMANVFRQOEYVAILDALAEYHNMSENVTVEKYLKLDINSLTDIYI	60
QY	61	NTYKSGRNLKALKKPKYELTMEVLELKNLSLTPVEKNLHFIWIGGQINDTAINYNQWKD	120
DB	61	DIYKSGRNLKALKKPKYELTMEVLELKNLSLTPVEKNLHFIWIGGQINDTAINYNQWKD	120
QY	121	VNSDYTVKPYVDSNAPLINTLTKTIVTESATNTNLESFRENLDNDPEPYNKFVKRMELIY	180
DB	121	VNSDYVNVVYDSNAPLINTLTKTIVTESAINDTNLESFRENLDNDPRDYNKFVKRMELIY	180
QY	181	DKQKHFDYKSOIENPEFIIDNIITKYLSEYKDEALNKYIYESLNKLTANNNDI	240
DB	181	DKQKNFINTYKQOREENPELLIDDIIVKTLSEYKDEALNKYIYESLNKLTANNNDI	240
QY	241	RNLKFEADRLVRLNQLVERNLAAASDILRISMLKEDGGVLDVLPQDPDLPKS	300
DB	241	RNFEEFNGESFNLYEQELVERNLAAASDILRISALKIIGMNYLDVMLPGIQQDLPFS	300
QY	301	INKPDSITNTSWMEMIKLEAIMKYEYIPGYTSKNFMDLDEEVQSFESALSKSKSEIF	360
DB	301	IEKPSVTVDFWEMTKLEAIMKYEYIPGYTSSEHFMDEEVQSFESVLSKSKSEIF	360
QY	361	LPDLDDIKVPLEKIAFANNVINOALISLKSQSDYVINOIKNRYKILNDNLAPSINE	420
DB	361	SSLGDEASPLEKIAFANNVINOALISLKSQSDYVINOIKNRYKILNLSNLPATSE	420
QY	421	GTDFNTMTKIFSDKLASIGNEDNMFMWIKITNLYKVGFPADVRSTINLSGPGVYTCAYQD	480
DB	421	DNDFTNTTTFIDSINVAEANADNGRFMELGKLVRFPPDVKTITINLSGPEAYAAAYQD	480
QY	481	LMFKDQNSNIHLPELARNPEPKTKISQLEQETISLWSNQARAKSQFEYKKGYPE	540
DB	481	LLMFKEGSMNIHLIEADLENFEISKTNIQSQTEQENASLWSPDARAKAQFEYKRYNYPE	540
QY	541	GALGEDNDLPAQNTVLDKYYSKTLSSMKTENKEYIHYIYVLOQDKISYEASCNLPSK	600
DB	541	GSIGEDNDLDFSQNIYVDKEYLEKLSLARSSERYIHYIYVLOQDKISYEAAACNLFAK	600
QY	601	DPYSSILYQKNTBEGSETAYVYVADAEIKEDKIRIPYQISNKRNIKLTFIGHKGSEFNT	660
DB	601	TPYDVLVFKNIEDSIAYYNPGDGEIQEIDKYKIPSIISDRPKILTFIGHKGDEFNT	660
QY	661	DTFANLDVDSLSSEIETILNLAADISPKYIEINLGCNMFYSYIYAETYPGKLLKTK	720
DB	661	DIFAGDVDSLSIEIAADLAKEDISPKSIEINLGCNMFYSINVEETYPGKLLKVK	720
QY	721	DRVSELMPSISQDSITVSNQYEVRIENEGKEILDHSGKWINKEESIHKDISKEYISF	780
DB	721	DKISELMPSISQDSITVSNQYEVRIENEGKEILDHSGKWINKEESIHKDISKEYISF	780
QY	781	NPKENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR	840

RESULT 2

US-10-011-366-6  
 ; Sequence 6, Application US/10011366  
 ; Publication No. US20030054493A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williams, James A.  
 ; Kink, John A.  
 ; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
 ; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

DB	781	NPKENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVMLTECEINVASNIDRQIVEGR	840
QY	841	IEEAKNLSDSINYIKNEFKLIESDSLYDLKHQGLDDSHFIFEDISKTENGFRIRF	900
DB	841	IEEAKNLSDSINYIKNEFKLIESDSLYDLKHQGLDDSHFIFEDISKTENGFRIRF	900
QY	901	INKETGNSIFETEKEIPEYATHISKEISNIKDTIFDNVNGKLVKNLDAAEHVNTLN	960
DB	901	INKETGESIFVETEKTIPEYANHITEEISKIGTIFDTVNGKLVKNLDTTHEVNTLN	960
QY	961	RAFFIQSILIEYNTTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVELVSTALDEITD	1020
DB	961	RAFFIQSILIEYNTTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVELVSTALDEITD	1020
QY	1021	LLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAITV	1080
DB	1021	LLPTLSEGLPIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAITV	1080
QY	1081	SALGTASGFSILLVPLAGISAGIPSLVNNELTLODKATKVIDYFKHISLAETEGAFITLD	1140
DB	1081	SALGTASGFSILLVPLAGISAGIPSLVNNELTLODKATKVIDYFKHISLAETEGAFITLD	1140
QY	1141	DKIIMPQDDLVLSEIDFNNSITLKGCEIWRAGSGSGHTLTDIDHFFSSPSITYRKPMWL	1200
DB	1141	DKIIMPQDDLVLSEIDFNNSITLKGCEIWRAGSGSGHTLTDIDHFFSSPSITYRKPMWL	1200
QY	1201	SIYDVNTKKEIKDPSKOLMVLPNAPNRFVGEWGTGFRSLDNDGTYKLLDRIDHVEG	1260
DB	1201	SIYDVNTKKEIKDPSKOLMVLPNAPNRFVGEWGTGFRSLDNDGTYKLLDRIDHVEG	1260
QY	1261	OPYWYFAPAFADALITLKPVEDTNRLNDGNTSRFIVPVTTEQRKNLSYFYSGG	1320
DB	1261	OPYWYFAPAFADALITLKPVEDTNRLNDGNTSRFIVPVTTEQRKNLSYFYSGG	1320
QY	1321	GSYLSLSPYNNIDNLVENDTWVIDVNVVKNITTESDEIQKGLIENILSKLIENIDN	1380
DB	1321	GSYLSLSPYNNIDNLVENDTWVIDVNVVKNITTESDEIQKGLIENILSKLIENIDN	1380
QY	1381	KIILNNHTINFGDINESNRFISLTFSTILEDINIIIEIDLVSYSKYLISGCMKLIENS	1440
DB	1381	KIILNSHINFSGEYVNGSFGVSLTFSTILEGINAIEVDLLSKYSKYLISGELKILMLNS	1440
QY	1441	SDIOQIDHIGFNGHQKVIYFVSYDNETKNGRFDYKSGELFTAEFSNESIIRNIYMP	1500
DB	1441	NHIOQIDYIGFNSLQKNIPIYSFVDSGKNGFNGSTGKGLFVSELPDVVVLISKVYMD	1500
QY	1501	DSNNLFYSSKDLKDIRINKGDVLLICNYFKODMKVLSLFTIEDTNTIKLNGVYLDEN	1560
DB	1501	DSKPSFGYSSNLLKDVITKDNVNLGYLKDIIKISLSITLQDEKTIKLSVHLDSE	1560
QY	1561	GVAQILKFMNNAKSALNTSNLSMNPESINIKNIPYNNLDNPIEFILDTNFIISGNSIG	1620
DB	1561	GVAELIKFMNR-KGNTNTSDLSMFSLESMNKSIFVNFQSNIKFILDFANFIISGTTISG	1619
QY	1621	QFELICDKDKNIQPIYFINKETSYTYLVGNRNLIVPEPSVHLDSDSGNISSTVINFOSK	1680
DB	1620	QFELICDKDKNIQPIYFINKETSYTYLVGNRNLIVPEPSVHLDSDSGNISSTVINFOSK	1679
QY	1681	YLYGIDRVNVKVIAPNLYT 1700	
DB	1680	YLYGIDRVNVKVIAPNLYT 1699	

DISEASE  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Medlen & Carroll  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/011,366  
 FILING DATE: 16-NO. US20030054493A1-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/957,310  
 FILING DATE: 23-OCT-1997  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 24-OCT-1994  
 APPLICATION NUMBER: US 08/161,907  
 FILING DATE: 02-DEC-1993  
 APPLICATION NUMBER: US 07/985,321  
 FILING DATE: 04-DEC-1992  
 APPLICATION NUMBER: US 07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ingollia, Diane E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPHD-01121  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2710 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-10-011-366-6  
 Query Match 51.4%; Score 4458; DB 15; Length 2710;  
 Best Local Similarity 50.7%; Pred. No. 2.2e-270;  
 Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY 1 MNLVNAQKQVYKFRIOEYVAILNALBEEYHNMSSSVVEKYLKLDINNLTDNYL 60  
 DB 1 MSLISKEELIKLAY-SIRPRENEYKILTNDYENKLTNNENKYLQKLKLNESIDVPM 59  
 QY 61 NTYKSGRNKALKKFKKEYTMEYLELKNNSLTPVERKNLHFIWIGQINDTAINYQWKD 120  
 DB 60 NKYKTSRRNALSNLKKDKILKEVILIKNSNTSPVERKNLHFWIGVSDIALEYIKQWAD 119  
 QY 121 VNSDYTVKVFYDSNAFLNLTWLTIVTESATNTLESFRENLDNPDYDNKFKRKRMEIY 180  
 DB 120 INAEYNIKLWYDSEAFVNLTKKAIYESSSTTEALQLEEEIQNPQDNMKFKYKRMET 179  
 QY 181 DKQKHFIDYKQIEENPEFIDNIITKYLNSYKDLKALNKYIEESLNKITTANNNDI 240  
 DB 180 DKQKRFINTYKQINQKPTVPTIDDIILKHLVSEYNRDETVELSYRNSLRKNSNHGIDI 239  
 QY 241 RNLEKFAEDLVRLNQLVERWNLAASDIARISMLKEDGGVYLDVILPGIQPDLPKS 300  
 DB 240 RANSLFTEQELNIYQELLNRGNLAASDIARLALKNFQGVYLDVDMLPGLHSLDKPT 299  
 QY 301 INKPSITNTWEMIKLEAIMKYKEYIPGYTSKNFMDLBEVQVSFESALSSKDSKETP 360  
 DB 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNTSENFDKLDQQLKDNFKLIIESKSEKSEIF 359

QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420  
 DB 360 SKLENLNSDLEIKIAFALSGVINQALISKQGSYLTNLVIEQVKNRYQFNLQHLNPAIES 419  
 QY 421 GTDFNTWKIFSDKLASISNEDNMFMKITYILVKVGFADPVRSTINLSGPGVYTCAYOD 480  
 DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISLSPGAYASAYD 479  
 QY 481 LLMFKDNSTNIHLPELRNFPPFKTKISQLTQEITSLWSFNQARAKSQFEYKKGYP 540  
 DB 480 FINLQENTIEKTLKASDLIEFKFEPENNLSQLTQEINLSLWSDQASAKYQFERYDYTG 539  
 QY 541 GALGEDDNLDPQNTVLQKDY-VSKKILSS--MKTRNKYIHYIYVQLQDGKISYEASCNL 597  
 DB 540 GSLSEDNQGVFNKNTALDKYLLNNKIPSNVVEEAGSKNYVHYIYVQLQDDISYEATCNL 599  
 QY 598 FSKDPYSYLQKNIEGSETAYYYVAD--ARIKEIDKYRIYQISNKNENIKLTFIGHGK 655  
 DB 600 FSKNPKNSIIQORN--NESAKSYFLSDGDSILELNKYRIPERLKNKKEVKVTFIGHGK 657  
 QY 656 SEFNTDTTANLDDVDSLSSEIETILNLAKADISPKYIEINLLGNMFSYSIYAEETYPGKL 715  
 DB 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLDISPKVVEVNNLLGNMFSYDFNVEETYPGKL 717  
 QY 716 LLKIKDRVSELMPISQDSITVSANQYEVRIINEEGKREILDHSGKWINKEESIIDKISSK 775  
 DB 718 LLASINDKITSLPDVNKNSITIGANQYEVRIINESEGRKELLAHSGKWINKEEAIMSDLSSK 777  
 QY 776 EYISFNPKENKIIVKSKYLHELSTLLOEIRNNANSDDILEKVMLTECEINVASNIDRQ 835  
 DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFINLNKLAIESSIGDY 837  
 QY 836 IVEGRIEERAKNLTSDSINVYKNEFKLIESISPSLDLKHQNGDLDHSHFISPEDISKTENG 895  
 DB 838 IYKLEPVKNIIRNSIDDLIDEPNLENSDELTELKLANLDEKYLISFEDISKNST 897  
 QY 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIDKTIIDNVNGKLVKKNVLDAAHE 955  
 DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNIQLDHTSQ 957  
 QY 956 VNTLNSAPPIQSLIENYNTKESLSNLSVAMKVQVTAQLPSTGLNTITDASKVVELVSTAL 1015  
 DB 958 VNTLNAAPFIQSLIDYSSNKOVDNLSVSKVQLYAQLPSTGLNTIYDYSIQLVNLISNAV 1017  
 QY 1016 DEFIDLLPTLSGLPIIATIIDGVISGAAIKELSETNDPLLRQETEAKEIGIMAVNLTAAS 1075  
 DB 1018 NDTIINVLPTITGPIVSTILDGINLGAAILKELDEHDPLLKELEAKGVGLAINLSLSI 1077  
 QY 1076 TAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGA 1135  
 DB 1078 AATVASIVGIGAEVITFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSKSKYP 1137  
 QY 1136 FTLLDDKIIMPODDLVLSEIDPNNSITLGKEIWRABEGSGHTLTDIDHFPSSPSITY 1195  
 DB 1138 LKTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGGSGHTVTGNDHFPSSPSISS 1197  
 QY 1196 RKPWLSIYDLNKKKEDIDFSKDLVLPNAPRVFCYENGWTPGERSLNDGDKTLDRIR 1255  
 DB 1198 HIPSISYSAIGIETENLDFSKKIMLPNAPSRFVWMTGAVPGLRSLNDGDKTLDRIR 1257  
 QY 1256 DHYEQVYRYFAFTADALITKLPRYEDTVNRINLDGNTRFVIVPVIETQIRKNLSYS 1315  
 DB 1258 DLYPKFWRFYAFP-DYAITTLKPYVEDTNKIKLDDKTRAFIMPTITITNEIRNKLVS 1316  
 QY 1316 FYSGGGSYLSLSPYNNMIDNLVNDTVVDVNVKNITTESDEIQGELIENTLSKL 1375  
 DB 1317 FDGAGGTSLLSYSPISNTINLSKDDLWIFNIDNEVEISIENTGKIKGLIKDVLSKI 1376  
 QY 1376 NIEDNKILNNTINFYGDINESNFISLTFSILEDINIIIDLVSKYKILLSGNCKM 1435  
 DB 1377 DINKKLIIGNOTIDFSGDIDNKDRIYIFLTCLDDKISLIIEINLVAKSYLLSGDKNY 1436  
 QY 1436 LIENSDDIQKIDHIDGNGHQYIFYSYID-NETKINGFIDYSKKEGLFTAESNESII 1494

Db 1437 LISLSNTTEKINTGLD---SKNTAYNTDESNNKYFCAI-----SKTSQKSI 1483  
QY 1495 RNIYMPDSNNL-----FIYSSKDL--KDIRINKGDVKKLLIGNYFKD---DMKVSLS 1541  
Db 1484 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVFMKDINTITGKYVDNNTDKSIDFS 1541  
QY 1542 FRIEDNTIKLGVLDENGVAQILKFMNNAKSALNTSNLNFLESINIKNIFYNLDP 1601  
Db 1542 ISLVSKNQVKNGLYNLSEVSYDFVKNSDGHNTSNFMNLFIDNISFWKLGFGE--- 1598  
QY 1602 NIEFIDTNTFISGNSIQOFELICDKDKNIOPYFINFKIKETSYTYLVGNRQNLIVEPS 1661  
Db 1599 NINFVIDKYFTLVGKTNLGVFEICDNNKNIDIFYGEWKTSSKSTIFSGNGRNVVEPI 1658  
QY 1662 YHLDGSGNSTVINPFSQKLYGIDRYVKNVJIAPNLYT 1700  
Db 1659 YN-PDTGEDISTSLDPSYBPLYGIDRYINKVLIAPDLYT 1696

## RESULT 3

US-10-114-170-257  
; Sequence 257, Application US/10114170  
; Publication No. US20030023075A1  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; Burland, Valerie  
; Perna, Nicole T.  
; Plunkett, Guy  
; Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI

COUNTRY: US

ZIP: 53701-2113

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170  
FILING DATE: 01-Apr-2002  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702  
FILING DATE: 03-DEC-1999  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998

## ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386

## REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166

## INFORMATION FOR SEQ ID NO: 257:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3169 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear

## MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 257:

US-10-114-170-257

Query Match 5.8%; Score 501.5; DB 15; Length 3169;

Best Local Similarity 20.1%; Pred. No. 4.6e-22;

Matches 421; Conservative 331; Mismatches 723; Indels 619; Gaps 104;

QY 14 YVKPRI---QEDEYVAIINAL--BEYHNMSSESSVVEKYLKLDKINLLTDNLYNTYKSGR 68  
Db 160 YIKIRKTRGAEDQTTTQSLLIINELLNGVDRNTI--PEKISELNDIIHSHENQIKNSR 218  
QY 69 N--KALKFKFEYLWVEVLEKNS-----LTPVEKNLHFTWIGQ 106  
Db 219 KGIETLYKQGLSSLLNDNKGKQSDNASKIINLLGIEYQSHKVDIEPIHAWVAGA 278  
QY 107 INDPAINYQKVDNSDYTVKVVDSNAP-----LINTUKT----- 144  
Db 279 PPDNTFSYITAFNTYKDYTYLLWDPNAPGAAPSGILKNIAMYAIIMRLRTRTNPLAE 338  
QY 145 -----IVESATNTLE--SFRENLDPEFY-----NKEYRKEWELIYDKQKFI 187  
Db 339 EMNEVILKIQIONETIEFKETRERLEKENRYKSLTSETREKFNVFLESIMGMQDNYF 398  
QY 188 DYYKSQIBENPEFI-----IDNIIK--TYLSNEYSKDLEALNKYTEESLNKITANGN-- 238  
Db 399 TYCISNGISNTDDISRDLFTNLVLSPEVQDNFKSTVEKNKRDIDLLKNTISQKFGDRP 458  
QY 239 ---DIRNLEKFADEDLVRLNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQP 295  
Db 459 QLRDINTLESFKKPDYFYQOEMLLRNWYAAASQVRAINILKEYGGYTTDTDLPAYSD 518  
QY 296 DLPKSINKPDSITNTSWEMIKLEAIMKY--EYIPG--YTSKNPMDLDEEVQSFESALS 351  
Db 519 KVSQIINE-KSDDKRFEDLKRRIISISILSLNGEKYKSIKH-DGLDETTLNQNLNL- 575  
QY 352 SKSKSEIFLPDDIKVSPLEVKIAFANNSVI-----NOALISLKDSSVC 395  
Db 576 SEIEK-----LTIDDY-FKPVETKVVDRDFKFKYQKWTENTWIRGNMFLTHKSKC 630  
QY 396 SDLVNQIKRYKILNDNLNPSINEGDTFTMTKIFSDKLASINE-----DNM 444  
Db 631 IDFILSGQKQYLEL-QRIRDNISYNNLFYTT-----EDLKSINVAIGGIPAKKYLEHG 684  
QY 445 MEMIKITNLYKVGAPDVRSTINLSGP----- 471  
Db 685 LF-----SEYRODGTIPYVVSTLNSIGSPDMIMQMKYKYSIGRGEVHIKONKUSDVNL 740  
QY 472 GVTYGAYQDLLMF---KDNSTNIHILLEPE-----LRNFEFPKTKISOLT---EQEITS 518  
Db 741 GYVASSNKDNKSFNWLNPVSVGINDIITDDESSWAVRNNDINKILFEKINCHVPEKLPTS 800  
QY 519 LWSFNQARAKSQFBEYKGYFEGALGEDDNLDFQAQNTVLDKDYVSKYLSSMKTRNKEYI 578  
Db 801 LY-----YEIDSRSPFGW---DNKSIKHVTEINKDLI--KDNILLTSSNIDV 844  
QY 579 HYIVQLOGDKISYEASCNLFKDPYSSITLYOKNEGSETAYYYVYVADAEIKEDKYRIPY 638  
Db 845 KLLIKL--DRELYAISKI-----DNPLAURISRTLQQLANVYVTSNTFPENTINFIY 896  
QY 639 QISNKRN-----IKLTFIGHGKSEFNTDTTFANLDVDS-----LSSEIET 677  
Db 897 DFKKQDDLLSAIKL-----FSRNDADTKIIVWYNSVMKXVFLREVISC 942  
QY 678 IINLAKADISPKYIEINLLGCNMFYSYIYAETVPGKLLKIKORVSELMPSISODSI-- 735  
Db 943 VLRSKKVD---SYINEN-----KKNLSEKDA--GALRDYAKLAKMKELFSLMLDDGYKK 990  
QY 736 TVSANQYEVRIINEGKREILDHSGKWINKESIIDISKEYISFNP-----K 783  
Db 991 IITTNAY---IKERDKL-----SGIYNIENSIISGHESFDIIRSNQHEWGLDSTVEQFK 1042  
QY 784 ENKIIVKSKYLHELSTLQEIIRNNANSDDIDLEKKVMLTECEINVAS-----NIDRQIVE 838  
Db 1043 KFEFYVKSE-LSSAKSIIFDDIKNK-YITDPETKRNVLVHQLDSDIKERIAFLDISHYAP 1100  
QY 839 GRIBAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNGL 878  
Db 1101 GSLLEKQLS---GYVPSDINIIEAYLLASVGVSHGVVYPAPSKLLELRHTK 1156



Db 1141 REKKDKNIKKRYKKNKFNPNVSLNINCNFSLKCLGNMRHNNLSLFKNLTTLTKTGEIELKL 1200  
QY 575 KEYIHVIV-----QLOGDKISYEASC-----NLFS--KDPY 603  
Db 1201 KWLHLVKNWFYKKGKMKYIKRKLKNNKKIYAYICIGDFSNCEYEHINNYLFKLKNFF 1260  
QY 604 SSI-----LYOKNIEGSETAY-----YVADAIEKEIDKRYIPYOI 640  
Db 1261 DNINNEFYIYLPKSRFLYKLNLSFLSYYPNVKSGFLHY-----IRNLRELIKSHL 1315  
QY 641 SNKRNIKLTFIHKSEFN-----TDTFANLVD--SLSSEIETILNAKADISPKYIEN 694  
Db 1316 NDNHFLNQMPKTSKSLYIFADSYKSLQVDRDIFMTIITVIRYVYLVYFYSKEPK 1375  
QY 695 LLGCNMFYSIYAEETYPGKLL-LKIKDRV-----SELMPSIQDSITV 737  
Db 1376 LARKNIFYQIFQENQMGVYLSVRDKRVENIKWYLSNMKKINHDEILESKNSSINI 1435  
QY 738 SANQVEVRINECKREILDHSGKWINKESI-K-----DISKEYISENPKENKII 788  
Db 1436 NKNFMICITHHQDTE---EKGNTQKKEKHDYIGPIYNNSPDSTTTTHSSNNYKGNH 1492  
QY 789 VKSKY-----LHELSTLQE-----IRNANSDDIDLEKVMLTECEINVASNDRQI 836  
Db 1493 VSGDYKNDGLLHKGNMNECYVKDIKCNMNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1552  
QY 837 VEGRIEAKNLTSDSI---NYIKNEFK--LISISDSLYDLKHQGLDP---SHFISPED 888  
Db 1553 I---IKYHKTIDTDSKNHNTYKFNKFLNFDLKKIISNYIYGLPQGSLSNLCSLYAYLD 1609  
QY 889 ISKTENGPRIRFNKETSNIETETEKEIFSEVATHISKEISNIKOTIFDNVNGK---- 943  
Db 1610 --KNEEFQNLVSEKQINKNYFLANGT---CHYFNLSLILFIDDFLITLKNKNIKIF 1664  
QY 944 ----LVKKYNLDAAEVNTLSNAPFIQSII-----EYNTTKESLSNLSVAMKV 987  
Db 1665 KNLLLKKKI---WGSNINSKTKIKFIPLIYKNDLLIYNFQNKYQKKYKIKKXKIQSV 1721  
QY 988 ---QVYAOQFSTGLNTITDASKVELVSTALDETDL--PTLSEGLPIIA--TIIDGVS 1041  
Db 1722 RNRKRHNQVLANKKKHTSVQK-----DKINKYINLIHPTIQKNDVSLSSNINWFERI 1775  
QY 1042 GAAIKELSTNDPPLRQRETEAKIGIMAVNLTAASTAIVTSALGIASGPSILLVPLAGISA 1101  
Db 1776 --YIKESHKN-----SSIRT 1789  
QY 1102 GIP-SLVNNELIQQATKVIDYFKHISLAETEGFTLLDDKIIMPQDVLVSEIDFNNN 1160  
Db 1790 DIPNSVWDDIEYNQSD-----NNSYSTNNLYNNINMTQNG-----DNN 1829  
QY 1161 SITLGCETIWRAGGSGHLLTDIDHFFSPSITVRKPLWSIYDVNLINKEIKIDFSKOLM 1220  
Db 1830 NVNIFK-----HYQNDSDFOCFNSN-----NLYIEKIDKENNI----- 1861  
QY 1221 VLPNAPNRFVGYEMGWTPGFRSLDNDGTLKLDRIHRYBGQFYWRFYAFIADALITKLKP 1280  
Db 1862 ---SQINRKLCKRNFTKSRKINTLYLQIDKV-----IKLKCKK 1900  
QY 1281 RYEDTNVRINLGNTRSFIVPVITTEQIRKNLSYSGFYGGSGYSLSLSPYNNMI-DLNLV 1339  
Db 1901 KYIKHKIKMKYMNQNF-----KKLKKLQKF---HNASFELKINKINKINRLNKL 1949  
QY 1340 ENDTWIDVDNVVNKNTIIESDIQKGLIENILSKLINIEDNKIILNNHTINIFYGIDNESN 1399  
Db 1950 KKR-----KHSI-----NITPVTSTB---WLANSYTFDP---INNS- 1980  
QY 1400 RFISLTFSLIEDNIIIEIDLVSYSKILLSGNCKMLIENSDDIOQKIDHIGFNGEHOKY 1459  
Db 1981 -----IQSTYP---WKNKCDATIRNHLHNVI-----IDKKNKY 2014  
QY 1460 IFYSYIDNETKYNPFIDYSKSEGLFTAEPFSNESIIRNIYMPDSNNLFYSSKDLDIRII 1519  
Db 2015 FMKNLVENRIVRNIISKQKQCSLYKNK-----QNVYFCYKNNFSLKSSILKFIICI 2067

QY 1520 NKGDVKLLIGNYFKDDMKVSLSFTIEDTNTIKLNGVYLDENGVAQILKFMNNAKSALNTS 1579  
Db 2088 -----KTLKKQFNFT-----NSTY-----NTRFIL--- 2088  
QY 1580 NSLNMFLSINIKNIFNNLDPNIEFILDTPFIISGNSISGOFELICDKDKNIQYFINF 1639  
Db 2089 -FLISYMNKMLIKN-----KLLKFV--KLFILQTAIEAFYARIFNQQDSFYCLOHF 2138  
QY 1640 -KIKETSYTYL-VGNRQNLIVPEPSYHLDDSGNISSTVINFSQKYLIDRYVNVKY 1692  
Db 2139 RKIKRKLINKYKIGHKNKLLREFFP-----LENFIKKELYNWPYMFKI 2182  
RESULT 5  
US-10-087-464-10  
; Sequence 10, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chieshti, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof  
; FILE REFERENCE: S1237/7019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; PRIORITY FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 10  
; LENGTH: 1639  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-10

Query Match 3.6%; Score 309; DB 15; Length 1639;  
Best Local Similarity 19.5%; Pred. No. 2e-10;  
Matches 351; Conservative 280; Mismatches 554; Indels 618; Gaps 94;  
QY 36 NMSSSVVVEKYLKLDINNLTNDYNTYKSGRNKALKKFKKEYLTMEVLELKNNSLTPE 95  
Db 112 NPSNSSSDAKSVADLKHVRNLLT-----IKELYPQFLDTNHNMLTLC 159  
QY 96 KNLHFIWIGGOINDTAINY-INQMKVNSDYTVKPYDSNAFLINTLKKTTIVESATNNTL 154  
Db 160 -NIH-----GPKYLDIGYEEIN-----ELLYKLN-FYDPLLRKLNDCANDYC 201  
QY 155 E-SFRENLDPEFDYK-----FYKRMELIYDKQKHFDIYDKSQIEENPEFIIDNIITY 209  
Db 202 QIPFNLKIIRANELVDLKKLVFGYRKPLDNIKDNVGMEDYIKKN----- 245  
QY 210 LSNEYSKDLKALNKYIEBSLNKITANNNGDNRNLEK-----FADEDLVRLYNQELVERN 265  
Db 246 -----KKTIEINELIEES-KTIDKKNATKEEKKLYQAYDL-SIYNKLEEAHNL 298  
QY 266 AAASDILRISMLKEDGGVYLDVILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAIMKYK 324  
Db 299 ISVLE-KRIDTLKKNENIKELDKINEIKNPPANSNGTPTNL-----LDKKNKIE 348  
QY 325 EYIPGYSKTFMDLDEEVQORSFESALSSKSKSEIFLPLDDIKVSPLEVK--IAPANNSV 382  
Db 349 EH-----EKEI-----KEIAKTIKFNIDSFLTDPLEYLEYREKNKI 386  
QY 383 INQALISLSDSVCSDLVLNQIKN--RYKILNDNLNPSINEGTDFNTMTKIFSDKLASISN 440  
Db 387 DISAKVETES-----TEPEYFNGVTYPLSYNDINNALNLSFG----- 427  
QY 441 EDNMFMMIKITNYLKVGFAPDVRSTINISGPGVYTGAYQDLLMFKNDSNIHLLPELRN 500  
Db 428 -----DUINPFDYT-----KEPSKNIY-TDNERKK 451







Db 291 INNK-----LNEINBEPYEDTHNKESSNKNNDKDEKKEE-----NNNITRY---IKN 337  
Qy 270 DILRISMLKEDGGV-----LDVILPGIQDLPKSNKPSITNTSWEMIKLEAIMKY 323  
Db 338 DQMSYNNINTNENEKDNASTIDETVIGKTFGGYVYVSNK-----NAACIKLKNNKIN 390  
Qy 324 KEYIPCVTSKNFMDLDEEVQSFESALSKSKSBIFLPLDDIKVSPLEVKIAFAN----- 379  
Db 391 -----GLLPKNKANLGDDE-----DNWDF-----EKQPVHVKILGINTKKN 429  
Qy 380 -----NSVI-----NOALIS-LKDSYCSDLVINQ-----IKNRYK 408  
Db 430 IYPLGNIIKYNENIKLSKEYSKGLITKVCDSYCFIKVLKNGSTGYLHKSKLFCMNDKEK 489  
Qy 409 ILANDLN-----PSI-----NEGTDFNTWK 429  
Db 490 KNDNDQNDPNNDPNYDQNYDQNYNPNYDHPNYDQNYDQNYNPNNDYSTQL 549  
Qy 430 IPSDKL-----ASISN-----EDNMFM-----KITNYL 454  
Db 550 YNSDNLQMDPIYKLOFTKIFNIWDIIDEILGTPQNDYKSNVILTI PRGSKTFKLLNVL 609  
Qy 455 KVGFPADVRSTNLGSGVYTGAYQDLLMFKDNSTNIHLLEPELNFPFPKTKISQLTEQ 514  
Db 610 NVLKENEEDINNIQYKGDYIS-----IDNNKNDNIIDSDINHHNNKKKK----- 656  
Qy 515 EITSLWSFNQARAKQFEBYKGYEGALGEDDNLDFQAQNTVLDKDYYSKILLSMKTEN 574  
Db 657 ---NLYDIQNNVHSPFNK-----HTEDEYLFNDHQENV-HTFYEKN 696  
Qy 575 KEYIHVQLQDKISYEASCNLFSPYSSILYOKNIEGSETAVYVYVADAEI----- 628  
Db 697 KKY-----KITDYKEN-----HKNN-----KSYLKKKKELFPNNKF 729  
Qy 629 KEIDK--YRIPYQISNK-----RNKLTFIGHGKSEFTDTTFANLVDVLSSEIE 676  
Db 730 KKIKNIVDPLTISLSMLSKTIKIPLASIKKYFIHENEKYNSSYKIN-----SEQIK 783  
Qy 677 TILNLAKADISPKYIEINLLGCMNPSYSIYAEETPGKLLKI-----KDRVSELMPISQ 732  
Db 784 RICQHKID-----CNV-----EQRDDNVVTVKNGTTCDCQSKVFNVTQ 823  
Qy 733 DSITVSANQYEVINEEGREILDSHGKWINKEESII---KDSSKEYISFNPKNKIIV 789  
Db 824 DRL-----KEGEQERVIVKEAKINDVMVQEQKTKKEKMDV-----862  
Qy 790 KSKYLHELSTLQEIERNNANSDDILEKKVMLTECEINV-----ASNIDQIVEGRIBEA 845  
Db 863 --QFTEEKDINVQHI-----NVQDMVDQD---MDVQDINVQDMVQNVQDI----- 905  
Qy 846 NLTSDSINVKNFEKLIBESISLDYLKHONGLDSDHPISFEDI-----SKTENGFRIRP 900  
Db 906 NTQDMQDVQNNISITLNTSKTSCQTDSDRDPAGDQNESLDEKDSMEKSKKKKKSKRK 965  
Qy 901 INKETGNSFIETEKIEPIEYATHISKEISNKTIDTFDNV--GKLKVKVNLDAAEVNTL 959  
Db 966 KMKDTN---LTLKSDSIQSKTLLDDKKNVVVTFIGHNKG-----1005  
Qy 960 NSAPFIQSLIEY--NTTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDE 1017  
Db 1006 -----TSLFDVICKTNEQKEYGLITQNIAPK---ATVRNFT-----FTLVDTPGHE 1051  
Qy 1018 TIDLPLTSEGPIATIIDGVSIGLAAIKELSETNDPLRQIEAKIGIMAVNLTAASTA 1077  
Db 1052 A--FMPMRSGVKISDLSILVSGDEGIOE-----QTV-----1083  
Qy 1078 IVTSALGIASGFSILLVPLAGISAGIPSL-----VNNELIQDKATKV-----IDYF 1124  
Db 1084 -----CIKLIKFNKII--IAITKVDIPNVVDRIINDLHYDITTELNGEIQVVECSIY 1138  
Qy 1125 KHISIAETAGFTLLDDKIIMPQDDLVLSEIDFNNNISITLKGCEIWAREGSGSHTITDDI 1184

Db 1139 KEESI-----DKLL-----DAIYLESEFLNLQTNPK-----KHEQAQGVVLDYSI 1179  
Qy 1185 -----DHFSSPSITYRKPMLSIYDVLNIIKKEKIDFSKDL--WVLPN 1224  
Db 1180 DKGIVSINLQNGVLNINDHIFTGSS-----YGVKILADHL--NKNVKSAYPS 1227  
Qy 1225 APNRVFGYEMGWTP-----GFRSLNDGTKLLDRIDHYEGQFYWRFYAFIADALITKLP 1280  
Db 1228 DPIKIIGYNKNSVPVAGDKFYVVENEA--LAKEIAEHKNKN-----LTWEINN 1274  
Qy 1281 -RYEDTVRINLDGNTRSFIVPVITTEIQIRKNLSYFSGSGSYLSLSPYNN--NIDLN 1337  
Db 1275 FTYDQTNM-----NRYKDFIIS-----REN-----KIGSSGI-LGENNLKNDIDGN 1315  
Qy 1338 LVENDTWIDVNVVNITIBESDEIOKGELIENILSKLNIENDKILNNHTINFYGDINE 1397  
Db 1316 MTRDDNMTFD-----DNMTSDNMTRDGNRTN-----DN-ITSDDNMSNDVKIKE 1361  
Qy 1398 SNRFISLTFSEIDINIIIBDLVSKYSKILLSGNCKML-IENSSDIOOKID---HIGFN 1453  
Db 1362 TKWYTN-----NKSQF---KODFLKIHLMNTNENVINMDPSTHIGKN 1400  
Qy 1454 GEHQXYIYFVIDNETKNGPIDYKKEGLFTAEFSNESIARN-----IYMPDS-----NNL 1505  
Db 1401 -----EIKTIYSNVIKCKQKQ-----SIEVLKNCMLKQKEDSICKIKNK 1441  
Qy 1506 FTYSSKDLIRIINKGDKVLLIGNYFKODMKVLSFTIEDTNTIKLNGVYLDENGVAQI 1565  
Db 1442 IYYAD-----IGNVTSSDIKYATSFNA-----TIIAFGVKLSND----- 1475  
Qy 1566 LKFMNNAKSALNTSMLNPLESINIKNIFNNLDNPIE 1604  
Db 1476 IKGSKNKGSKXNHN-----YPIIYSNVLYELIE-NVE 1507

## RESULT 8

US-10-032-585-7646  
; Sequence 7646, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7646  
; LENGTH: 1881  
; TYPE: PRT  
; ORGANISM: Candida albicans  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1881)..(1881)  
; OTHER INFORMATION: X-any amino acid  
US-10-032-585-7646

Query Match 3.3%; Score 284; DB 12; Length 1881;

Best Local Similarity 19.4%; Pred. No. 9.2e-09;

Matches 277; Conservative 231; Mismatches 451; Indels 468; Gaps 65;

Qy 13 VYVFPRIQDEYVAILNAALEYHN--MSSESVVVEKYLKLDIN-----NLTDNVLNTYKKS- 66

Db 697 VYV--NLVKDNFLRIKRAL--FHDPSSEPRGISYELPEELDTKVLNLTKE-LQTEKENA 751

Qy 67 -GRNKALKKFKREYLT-----MEV-LELKNNSITPVEKNLHPFWIGQINDTAINYNQWK 119

Db 752 ESNDELNEKEKTEKLTNLSTKLETKLEDEQELAKIQED-----HK 791

Qy 120 DVNSDYTVKVFVYDSNAF-----LINTLUKKTIVESATNTLESFRE 159

Db 792 SLNE-----KFLVTANSGLGKARTKESETISGPDQOELQALK---GNTSESTLKLQKE 844  
Qy 160 NLNDPEFDYK-----FYRKRMEI-----IYDKQHF--IDYKQIEENPEF 200  
Db 845 KJOSTQAKKULEGGINNTRDLFHUKKSBAETOIKOREREFKMLTYEFFENTKRYDEL 904  
Qy 201 IIDNIKTLYSNEYSKDLBALNKYIBESLNKITTANNNDIRNL-EKFADEDLRLYNQEL 259  
Db 905 QINNLNKS--NNEFKQINELSKIE---SLTEDNKNFAQLKEEKLRTDE---ENNEHL 955  
Qy 260 VERWNLAAASDLIRISMLKEDGGVYLDVLDILPGIQDLPFKSINKPDSITNTSWEMIKLEA 319  
Db 956 MDKLRASVA-----YNDLKAKSESE-----EE 979  
Qy 320 IMKYKEYIPGYTSKNFMDLDEEV-----QRSEFALSXSXSK-DSEIFLPLDD-----IKV 368  
Db 980 TVKAKEBELTTSK-IDNLEKELKEQSKKLEGOLOQITDSTWNEKFEDELKSIXK 1038  
Qy 369 SPLEVKIATFANNSVINOALISLKSYSCLVINOIKRYKILNDNLNPSINEGTFNFTM 428  
Db 1039 SNKE--ISSQSELIOKLEKTEKLOAKDEBDKLAETKSNIDNLNSEI---SSLQSKL 1093  
Qy 429 KIFSOKLASISNEDNMFMKITYNLKVGFPDVRSTINLSGPGVYTGAYQDLLMPKONS 488  
Db 1094 KEAESHSSTKDEHSSL-----S 1111  
Qy 489 TWHILLEPELRNPEFPKTKISOLTE-----QEITSLMSFNQARAKS 529  
Db 1112 ENLKKLKEEYEN-----TKTSMIAKLSAKIEBHKATDIEITKTHITDLOE-EHAKQKS 1165  
Qy 530 QEEYKKGYPFEGALGEBDDNLDPAQNTVLDKDVSKKILSSMKTRNKRYTHYVQLQGDKI 589  
Db 1166 QFESERNDI-----KSNLDRANKELSD---NREKLSNLE-KETELNKKLKTQOEKI 1213  
Qy 590 SYEASCNLSPKDPYSILYQKNIEGSETAYYYVADAIEKIDKVPYQISKNRNIKLT 649  
Db 1214 SDLETSVAISEDKSKSLKH-----DIEDLKREKIKLETTKENEETM 1255  
Qy 650 FIGHGKSEFNTDFANLD-----VDLSLSEIETI---LNLAKADISPKYIENLLGCNM 700  
Db 1256 FKKEQLOVNDCKKELEACLLKTETKEKENDLIRKLEAAKSD-----S 1300  
Qy 701 FSYSIYAETYPKGLLLKIKDRVSELMPISIQDSITVSANQVEVRINEGRKELIDHSGK 760  
Db 1301 -----HTERKKSLLLEDYKSE-----SEKNVILKNEQIEKLGEREKEVRDIQSQ 1347  
Qy 761 -----W-----INKEPSITKDISKBYISFNPKNKIIVK-----S 791  
Db 1348 LAAKTTDWEKIKTTLDKVLKESDLEK-TNKESVDTLKKVENLKEISLLEDQKDDTT 1406  
Qy 792 KYLHELSTLQIRNANSDDIDLEKVMLTCEINVASN-----IDRQIVE 838  
Db 1407 KY-KELAAQLETKTSNLDSTTMELEK-----TELELKKVNELETATSELTKLQDNQSLT 1461  
Qy 839 GRIEEA-----KNLTSDSINVINEPKLIBS-ISDSLYDLKHONGLDD 880  
Db 1462 EBEIKTKAALTSSKDLVCGNQKSELQDSLSKVSSELKNPKENKYNQETTSKLADIEBKQ 1521  
Qy 881 SHFIFGEDIKTENGFRIRFINKTGNISFIEETEKEIFSEYATHI-----S 926  
Db 1522 KEIVTLQ-----TELKDRISEVEKE--RAMLSENSETVIKEYSDKISLESKINSIKENS 1575  
Qy 927 KEI-----SNIKOTI-----FDNVGKLVKKYN-----LDAAEVNTLNSAFPIQSIL 969  
Db 1576 KEITTHNEOKTSLKQDIAKLSQDHSAQOQLDKEKNQNLKELKASLEKHNTESA---TSIE 1632  
Qy 970 EYNTTKESLNSLVAMKVOVYLAQLFSTGLNTITDASKVVVELYSTALDETIDLLPTLSEGL 1029  
Db 1633 EKXNQIKELSEIYKSLKTEL-----KTSGDALKQSQOKEYKTLTKTNSDTSKLEKQLEEL 1687  
Qy 1030 PIATIIDGVSIGAAIKELSETNDPLRQEIIRAKTIGIMAVNITAASTAVTISALGIASGF 1089

Db 1688 -----EKVKSQDLQATDEKLGITEREIALKSELETNVKNSGLSTTS----- 1727  
Qy 1090 STILLVPLAGISAGIPSLVNNELILO-----DRATKVVDYF-KHISLAETEGFTLLDDKII 1144  
Db 1728 -----ELAALTKTVKSLEKEEEOFLSGNKSKELEDIYQKHSIDISEKKAUT----- 1775  
Qy 1145 MPQDDLVLSEIDFNNSITLKGCEIWRAGGSGHTLTDIDHFFSSPSITTKPKWLSIYD 1204  
Db 1776 ---DELAKETKQFDDSKKKLLELE-----NDLTS-----TKKELETEKTQTSKFK 1817  
Qy 1205 VLNIKKEK--IDFSKDLMLVPLNAPNRVFGYEMGWTGPRSLDNDGTK 1249  
Db 1818 NLEERKDEIVKLNKLELLKN-----DNSGAK 1845  
RESULT 9  
US-09-801-368-56  
; Sequence 56, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801.368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 56  
; LENGTH: 2167  
; TYPE: PRN  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-56  
Query Match 3.1%; Score 268.5; DB 10; Length 2167;  
Best Local Similarity 18.3%; Pred. No. 1.1e-07;  
Matches 341; Conservative 276; Mismatches 610; Indels 633; Gaps 83;  
Qy 53 NNLTDNYLNTYKSGRNKALKKFKLEYLTMEVLELKNLSLTP-----VEKNLHPIW 102  
Db 141 SNAQDETPIAYKKSAAKSD-----SLLPSRSSLSPPQSRCSGTGTTLEKSLNTSG 190  
Qy 103 I-----GGQINDTAINYNQWQVNSDYTVKVFYDSNAFLINTLTKTIVES-ATN----- 151  
Db 191 ISNSGTTNNNSNNNDNEQQRNVHILNSENYYDTTVFKTGVNKSHTQVATNYSNMT 250  
Qy 152 ---NTLESFRENLDPEFDYN---KFY-----RKMELIY 180  
Db 251 APSSSSSSSSSQNLNDAYSRRNRSRFGYNGDSSSLKNDSSSTTATNSGNDVASARSSHAI 310  
Qy 181 DKQKHFDY--YKSQIEENPEFIIDNIITKLYSNEYSKDLKALNKYIEESLNKITANNNGN 238  
Db 311 DPQMLVPDYRLYRAQ-----LKGCVNLNLYKSLGSLNFKFPDPTL---PASNSS 355  
Qy 239 DIRNLEKFADEDLRLYNQELVERWNLAAASDLIRISMLKEDGG-----VYLDVLDILPGI 293  
Db 356 -----IANEN-----HQKKQQTNOAQAEALH---QKQSFQMGEPITLIDLKYLSEV 400







```
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-932-183A-2

Query Match          3.0%; Score 256; DB 10; Length 2285;
Best Local Similarity 17.8%; Pred. No. 7e-07;
Matches 313; Conservative 279; Mismatches 595; Indels 568; Gaps 76;

QY 142 KKTIVSATNN-----TLESFREN---LNDPEFYNNKFKYKMEI---Y 180
DB 781 QQTINVEAITTKDSTDKLIQYKELQKVKSRLTSDEREYLVQVTQQLAQTFPALVKGY 840
QY 181 DKQKHFDYKSOIENPEFIIDNIIKTVLSNEYSKDLALNKYTESLNKITANNNDI 240
DB 841 DSGGNAILTKNKELEK---AIEN-TKEYLALKQETRDSAKTTFEDASKEIKSK-DEL 894
QY 241 RNLEKFADEDLVRLYNOELVERNLAA-----ASDILRISMLK-----EDG----- 281
DB 895 KQYKQIAD-----YNDKGRPKWDLIADDDYKVAADKAKQSMKQAQSDIESNAKVDS 948
QY 282 -----GVYLDVILPGIQDPLFKSINK-----PDSITNTSWEMIKLEAIMKYEYTP 328
DB 949 VLSIANAYSIDISNTLTKTSISDVVNKLKDDLDPEELEKFSLSGLKQE--KMQKALD 1006
QY 329 GYTSKNFMDLBEVQSFESALSSKSKSEIFLPLDDIKVSPLEVKIAPNNSVINQALI 388
DB 1007 SGDEKAFDNKADQLSLLET--YKSDSS-----IDVFKMS-----FDKAKQ 1046
QY 389 SLKSDYCSDLVINQIKRYKILNDNLNPSINEGTDFNTMTKIPFSDKLASISNEDNMFM 448
DB 1047 NIKD-----GDKSLSVKSEVCDLGETLAEGNAEDEFK-----1081
QY 449 KITNYLKVGFAPDVRSTINLSGGVYTGAYQDILL---MFKDNTNTHLLEPELRNPEFPK 505
DB 1082 KLKEALDANSVDDIKAAIKEMSDAMQFDSVQDVLNGDIFNNTKQDVAPLNDLLERMAEGK 1141
QY 506 TKISQLTEQEITSLWSPNQARAKSQPEEYKGYFEGALGEDDNLDPQNTVLDKDYVSKK 565
DB 1142 S-----ISANEA-----NTLIQKQ---KE 1157
QY 566 ILSSMKTRKEYTHYIVOLQGDKI SYEASCNLPFKDPYSSILYKQKNEGSETAYVVYVAD 625
DB 1158 LAQAI SIENG-----VVKINRDE-----VIKQKV-----1182
QY 626 AEIKEDKIRIPYQISKNRNIKLTFIGHGKSEFNTDTFANLDVDSLS-----673
DB 1183 -----KLDAYNDMVTYSNK--LMKTEVNNAIKTLNADT---LRIDSLLKLRKERKLDMSA 1233
QY 674 -----ELETILNLAKADISPKYTEINLGCNMFESIYAEETPGKLLKIKDRVSELM 728
DB 1234 ELSDLVEKSIINNADAKELKLEEKMLQPGGYSNSQI-----BAMQSVKSALESYIS 1286
QY 729 SISQDSITVSANO-----YEVNRNEEGK-----REILDHSGKWINK 764
DB 1287 ASEATSTQEMNKQALVEAGTSLNNTDQEKANEETKTSWYVVDKYKALEKVNAEIDK 1346
QY 765 EESIIKDIS--SKYEISFNPKNKII--VKSKYLHELSTLLQEIIRNNANSDDILEKVM 821
DB 1347 YNQVNDYPKYSQYRDAIKKEIKALQKKLMQEQAKLLKDQIKSGNITQYGI VTTSTS 1406
```

RESULT 13

US-09-726-949A-1

; Sequence 1, Application US/09726949A







Query Match	2.9%	Score 249;	DB 9;	Length 1706;
Best Local Similarity	19.5%;	Pred.No.1.2e-06;		
Matches 349;	Conservative 261;	Mismatches 650;	Indels 534;	Gaps 90;
Qy	95	EKNLHFHWIGGQINDTAINYNOKWNSDVTVKFVVDNAFLINTLTKTIVESATNTTL	154	
Db	111	ESNLF-----NSSVLOGTNQITGYVEDGTLSTLSTSD-----LQSGIHKNTASDKY	157	
Qy	155	ESFRENL-NDPEFDYNKPF-YRKRMEIYIDKQKHFIDYKYSQIEENPFIDNIITKTVLSN	212	

Db	158	ENVELTILKSTNGCKYKNFATSNKQDWTFSQK-----NALLRS	194
Qy	213	EYSKOLREALNKY--IBESLNKTIANNNGDIRNLKEF---ADEDLVRLYNQELVERWNLA	267
Db	195	EYQADYESLRFPSLLSGSLNHSGLNADILGTDKINGSAHKATLRI-GQDGIS-----TS	249
Qy	268	ASDILIRISMLKEDGGVYLDVDIILPGTOPDL-----FKSINKPDSITNTSWHEMIKLEAIMK	322
Db	250	ATTNLKCSLLVLLENELNAELG-LSGASMKLTTNGRFRERNAKPSLDGKA-ALTELSLSGA	307
Qy	323	YKEVIGPYTSKNF-----DMLDEBQVRSFESA-----LSSKSD--	355
Db	308	YQAMILGVDSKNLFNFKVSOEGLKLNDDMGSAEMKFDHTNSLNTAGLSLDFSSKLDNI	367
Qy	356	-----KSBIPLPL-----BDIKVSPLEKIAFANNSVINOALI-----SLKD	392
Db	368	YSSDKFYQFVNLQLOPYSVLTTLSNLDKYNALDL---TNNCKLARLEPLKLVAGNLKG	423
Qy	393	SYCSDLVINOIKRYKILNDLNLPSTN-----EGTDFNTWKIIFSDKLASI-----S	439
Db	424	AYQN---NEIKHIYAISSAALSASYKADTVAKVQGVESFHRLTNTDIAGLASAIDMSTNY	479
Qy	440	NEONMMFMIKITNVLKVGEP-----DVRSTIN--LSGPGVYTGAYQDULLMPK-----	485
Db	480	NSDSLHF-----SNVFRSVMAFFTWTIDAHTNGNGKJALMGEHTGQGLYKFPLLKAEPLAFT	535
Qy	486	-----DNSTNIHLLEPLRNFEPKTKISOLTQEITSLWSFNOARAKSOFEBYKKGYPE	540
Db	536	PSHDYKGSTSHHLVSRKSI5AALEHKV5ALLTFAEQGTW-----KLKTQFNNEYSODL	590
Qy	541	GALGEDDNL-----DFAQNTVLDKDY5SKKILSSMKTNRKBYIHYIYVOLQ-GDKIS--	590
Db	591	DAYNTKDKIGVELTGRTLADLTLDSPKVPLLS-----EPINIIDALEMRAVEKPE	643
Qy	591	YEASCNLFSDPYSSILYQKNI5EGSETAYYYYVADAEIKEIDKVRIPYQISN-KRNKILT	649
Db	644	QEFITVAFVKYDKNQDVHSINLFPFFLQBYF-----ERNQTIIVVLLENVORNULK--	694
Qy	650	FIGHKSEFNTDFANLDVDSLSSEIETILNLAKADISPKYIBINLLGCNMF5YSIYAAE	709
Db	695	-----HINIDQFVR-----KY-----R	706
Qy	710	TYPEKLLLTIKDRVS-----ELMP5ISGDSITV5ANQYEVRAIN5EGKREIILDHSGKWINKE	765
Db	707	AALGKLQPOANDYLSNFW5RQV5SHAKEKLTALTCKY--RITENDIQIALDDA-----	757
Qy	766	ESIIKDISKEYISFNPENKI-----YKSKY-LHELSLTLOEIRNANS5DIDL	815
Db	758	-----KINFNEKUSLOQTYMIQFQYIKDSYDLHDLKIAI-----ANIIDEII	800
Qy	816	EKKVMLTE---CEINVASNI-DROIVEGRIE---EAKNLTSDSINYIKNEPKLIES5DSL	869
Db	801	EKLKSLDEHYHIRVNLVKTTHDLHFIENIDFNKSGSSTASWIONVDTKIQRIOQEKL	860
Qy	870	YDLK-HONGLDDSHFISFEDISKTENGFRIRFINKETGNSFIFETEKEI5EYATHISKE	928
Db	861	QOLK8H1QNIDIQHLAG--KLKHQIERAIDVRVLDDQGTTISPERINDIL-EHVKGH--V	915
Qy	929	ISNIKD--TIFDNVNG-----KLVKKVNLDDAHEVNTLSNAFFIQSILIE-----YNTTK	975
Db	916	INLIGDPEVAEKINAFRAK5VH5ELIEREYVDQIQV-----LMDKLVELLAHOYKUKETI	968
Qy	976	ESLSNLSVAMKVQVYAOLFSTGLNTITDASKVVELYSTALDEITDILPTLSEGLPIIATI	1035
Db	969	QKUSNVLQVKKIKNDYF-----KLVGFI-----DDAVKKLNEUS-----FKTF	1006
Qy	1036	IDGVS---LGAAIKEL-----SETNDPLLRQEIEAKIG--IMAVNLTAASTAI-----	1078
Db	1007	IEDVNFELDWLIIKKLSFDYHQFVDEINDKI--REVYQRLNGEIQALELPQKAEALKFL	1064
Qy	1079	-----VTSALGIAGSFTILLVPLAGISAGISPLVNNELILOD-KATKVIDYFKH-----ISL	1129
Db	1065	EETKATVAVYLES-----LOOTKITILINLOALSASL	1099



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:27:53 ; Search time 23.75 Seconds  
(without alignments)  
3028.567 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1700  
Perfect score: 8677  
Sequence: 1 MNLVNAQLQKVVYKPRIQ.....YLYGIDRVYVNVIIAPNLYT 1700

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6733.5	77.6	2366	1 US-08-480-604A-10	Sequence 10, Appl
2	6733.5	77.6	2366	2 US-08-405-496A-10	Sequence 10, Appl
3	6733.5	77.6	2366	3 US-08-915-136-10	Sequence 10, Appl
4	6733.5	77.6	2366	4 US-08-957-310-10	Sequence 10, Appl
5	6733.5	77.6	2366	4 US-10-011-366-10	Sequence 10, Appl
6	4458	51.4	2710	1 US-08-480-604A-6	Sequence 6, Appl
7	4458	51.4	2710	2 US-08-405-496A-6	Sequence 6, Appl
8	4458	51.4	2710	3 US-08-915-136-6	Sequence 6, Appl
9	4458	51.4	2710	4 US-08-957-310-6	Sequence 6, Appl
10	4458	51.4	2710	4 US-10-011-366-6	Sequence 6, Appl
11	501.5	5.8	3169	4 US-09-453-702B-257	Sequence 257, Appl
12	348	4.0	10182	4 US-09-134-001C-3159	Sequence 3159, Appl
13	310	3.6	2184	4 US-09-417-485D-6	Sequence 6, Appl
14	287	3.3	3433	4 US-09-091-501B-10	Sequence 10, Appl
15	281	3.2	3696	4 US-09-134-001C-5080	Sequence 5080, Appl
16	266.5	3.1	2662	4 US-09-595-684B-31	Sequence 31, Appl
17	261.5	3.0	3248	1 US-08-353-700-1	Sequence 1, Appl
18	261.5	3.0	3248	5 PCT-US95-16216-1	Sequence 1, Appl
19	261	3.0	2504	4 US-09-328-352-5821	Sequence 5821, Appl
20	257.5	3.0	3135	1 US-08-323-170B-2	Sequence 2, Appl
21	257.5	3.0	3135	4 US-08-954-441-2	Sequence 2, Appl
22	256	3.0	2285	4 US-09-308-375-2	Sequence 2, Appl
23	253	2.9	1296	1 US-08-480-604A-28	Sequence 28, Appl
24	253	2.9	1296	2 US-08-405-496A-28	Sequence 28, Appl
25	253	2.9	1296	3 US-08-915-136-28	Sequence 28, Appl
26	2391	2.9	2391	2 US-08-446-855A-2	Sequence 2, Appl
27	250	2.9	2391	3 US-09-150-741-2	Sequence 2, Appl

28	249	2.9	4536	4 US-09-180-422B-27	Sequence 27, Appl
29	249	2.9	4563	4 US-09-108-006C-1	Sequence 1, Appl
30	241	2.8	1786	3 US-08-973-462-8	Sequence 8, Appl
31	239.5	2.8	1169	4 US-09-255-829-20	Sequence 20, Appl
32	233.5	2.7	3878	4 US-09-914-259-11	Sequence 11, Appl
33	233	2.7	2522	3 US-09-251-645-13	Sequence 13, Appl
34	231.5	2.7	1346	1 US-08-471-033-23	Sequence 23, Appl
35	231.5	2.7	1346	2 US-08-471-044-23	Sequence 23, Appl
36	231.5	2.7	1346	2 US-08-463-483A-23	Sequence 23, Appl
37	231.5	2.7	1346	2 US-08-471-046A-23	Sequence 23, Appl
38	231.5	2.7	1346	2 US-08-470-566B-23	Sequence 23, Appl
39	231.5	2.7	1346	3 US-08-469-334-23	Sequence 23, Appl
40	231.5	2.7	1346	3 US-09-300-529-23	Sequence 23, Appl
41	230.5	2.7	3290	4 US-09-328-352-5486	Sequence 5486, Appl
42	226.5	2.6	1338	1 US-08-471-033-50	Sequence 50, Appl
43	226.5	2.6	1338	2 US-08-471-044-50	Sequence 50, Appl
44	226.5	2.6	1338	2 US-08-463-483A-50	Sequence 50, Appl
45	226.5	2.6	1338	2 US-08-471-046A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1  
US-08-480-604A-10  
; Sequence 10, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2366 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-480-604A-10

Query Match

Best Local Similarity 76.4%; Score 6733.5; DB 1; Length 2366;

Mismatches 190; Conservative 1299; Indels 1; Gaps 1;

QY 1 MNLVNAQOKVYKFRIOEYVAILNALBEYHNMSSSVVEKYKLKDINNTDNYL 60  
 DB 1 MSLVNRKQLEKMANVRFTQDEYVAILDALBEYHNMSENVTVEKYKLKDINSLTDYI 60  
 QY 61 NTYKSGRNKALKKPKKYLTMVELELKNNSLTPVEKNLHFINTGQINDTAINYNQWD 120  
 DB 61 DTYKSGRNKALKKPKKYLTMVELELKNNSLTPVEKNLHFINTGQINDTAINYNQWD 120  
 QY 121 VNSDYTVKPYDSNAPLINTLAKTIVESATNTNLTESFRENLDNPEFDYKPKRMEIYY 180  
 DB 121 VNSDYTVKPYDSNAPLINTLAKTIVESATNTNLTESFRENLDNPEFDYKPKRMEIYY 180  
 QY 181 DQKHPIYKQIENPFIIDNIITKTYLSNEYSKDLBALNKYIEESLNKITTANNNDI 240  
 DB 181 DQKNFINTYKQIENPFIIDNIITKTYLSNEYSKDLBALNKYIEESLNKITTANNNDI 240  
 QY 241 RNLEKPADLRLYNQELVERNLAAASDIIRISMLKEDGGYLDVLDVILPGIQDLPKS 300  
 DB 241 RNLEKPADLRLYNQELVERNLAAASDIIRISMLKEDGGYLDVLDVILPGIQDLPKS 300  
 QY 301 INKPSITNTSMEMIKLEAIMKYKEYIPGYTSKNFOMLDEEVQRFESALSSKSKSEIF 360  
 DB 301 INKPSITNTSMEMIKLEAIMKYKEYIPGYTSKNFOMLDEEVQRFESALSSKSKSEIF 360  
 QY 361 LPDDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420  
 DB 361 SSGLDMEASPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420  
 QY 421 GTDENTMTKIPSKLASISNEONMEMIKITNYLKVPADYRSTINLSGPGYVTCAYOD 480  
 DB 421 DNDFTNTTFTDISMAEADNGRFMBELGKYLVRGFFDPDKTINLSGPEAYAAAYOD 480  
 QY 481 LLMFKONSTNIHLLEPELRFNFPKTKISQLEQETISLWSFNQARAKSQPEEYKGYPE 540  
 DB 481 LLMFKONSTNIHLLEPELRFNFPKTKISQLEQETISLWSFNQARAKSQPEEYKGYPE 540  
 QY 541 GALGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKEYIHYIVOLQGDKISYASCNLFSK 600  
 DB 541 GSLGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKEYIHYIVOLQGDKISYASCNLFSK 600  
 QY 601 DYSSTLYOKNTEGSTAYVYVADAEIKEDIKRIPYOISKNRKIKLTFIGHGSEFNT 660  
 DB 601 TPVDSLVFQKNIEDSIAIYVNGDGEIQEIDIKYKIPSIISDRPKIKLTFIGHGSEFNT 660  
 QY 661 DTFANLDVDSLSSEIETILNLAKADISPKYIEINLILGCNMFYSIYAEETYPGKLLIK 720  
 DB 661 DTFANLDVDSLSSEIETILNLAKADISPKYIEINLILGCNMFYSIYAEETYPGKLLIK 720  
 QY 721 DRVSELMPSISQDSITVSANQVEVRINEEGKREILDHSGKWNKESIIKDISKEYISF 780  
 DB 721 DRVSELMPSISQDSITVSANQVEVRINEEGKREILDHSGKWNKESIIKDISKEYISF 780  
 QY 781 NPKENKIIYKSVKYLHSLTLQEIERNANSDDILEKVMLTECEINVASNIDRQIVEGR 840  
 DB 781 NPKENKIIYKSVKYLHSLTLQEIERNANSDDILEKVMLTECEINVASNIDRQIVEGR 840  
 QY 841 IBEAKNLTSDSINYIKNEFKLIESISDLYDLKHQNGLDDSHFISPEDISKTENGFRIRF 900

DB 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLQKNELEDSHFISFEDISETDEGFSIRF 900  
 QY 901 INKETGNSIFETEKEIEFSEYATHISKEISINIKOTIPDNVNGKLVKKNLDAHVEVTLN 960  
 DB 901 INKETGESIFVETEKTIIESEYANHITIEISIKGTIFDTVNGKLVKKNLDTTHEVNTLN 960  
 QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYLAQLRSTGLNTITDASKVELVSTALDEITD 1020  
 DB 961 AAFPIQSILIEYNTSSKESLSNLSVAMKVQVYLAQLRSTGLNTITDAAKVELVSTALDEITD 1020  
 QY 1021 LLPTLSEGLPIIATIIDGVSLGAAIKELSETNNDPLLRQIEIEAKIGIMAVNLTAASTAVT 1080  
 DB 1021 LLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQIEIEAKIGIMAVNLTAASTAVT 1080  
 QY 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELILQDKATKVIDYFKHISLAETEGEFTLLD 1140  
 DB 1081 SSGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVIDYFKHISLVETEVEGFTLLD 1140  
 QY 1141 DKIMPODDLVLSEIDFNNSITTLGKCEIWRAGGSGHTLTDIDHFFSSPSITYRKPLW 1200  
 DB 1141 DKIMPODDLVLSEIDFNNSITTLGKCEIWRAGGSGHTLTDIDHFFSSPSITYRKPLW 1200  
 QY 1201 SIYDVNLIKKEDIDFSDKDLMLPNAPNRVFGYEMGTGFRSLDNDGDKLDRIRDHYEG 1260  
 DB 1201 SIYDVNLIKKEDIDFSDKDLMLPNAPNRVFAWGTGTPGLRSLDNDGDKLDRIRDHYEG 1260  
 QY 1261 QFYRYFAFIADALITKLPYEDTNVINLDGNTRSPFIVPVITTEQTRKNLSYFYSGG 1320  
 DB 1261 EFYRYFAFIADALITKLPYEDTNVINLDGNTRSPFIVPVITTEQTRKNLSYFYSGG 1320  
 QY 1321 GSYLSLSPYNNIDNLVENDTVIDVDNVVKNITTESDEIOKGELENILSKLNIEDN 1380  
 DB 1321 GYALSLSQYNNIDNLVENDTVIDVDNVVKNITTESDEIOKGELENILSKLNIEDN 1380  
 QY 1381 KIILNNHTINFGYDINESNRIFSLTFSILEDINIIEIDLVSYSKYILLSGNCKMLIENS 1440  
 DB 1381 KIILNSHEINFGYDINESNRIFSLTFSILEGINAIEVDLLSKYSKYLLISGELKILMNS 1440  
 QY 1441 SDIOQKIDHIFNGEHQKIFYSYIDNETKNGFDIDYKKEGLFPAEBSNESIIRNIYWP 1500  
 DB 1441 NHIQOQIDYIGFNSLQKNIPYSFVDSGKENGFTNGSTKEGLFVSELDPVVVLIISKVYMD 1500  
 QY 1501 DSNLFIYSKDLKDIRIINKGDVKLLIGNYFKDKDMKVSLSFTIEDTNTIKLVYVLDEN 1560  
 DB 1501 DSKPSFGYNNKDKVKITKDNVNLITGYLKKDKIKLSLUTQDEKTIKUNSVHLDRES 1560  
 QY 1561 GVAQILKFMNNAKSALNTSNLIMNPLESINIKNIFNNLDNPIEFILDTNFIISGNSIG 1620  
 DB 1561 GVAELKFMNR-KGNTNTSDSLMSFLESNMNKSIFVNFQSNIKFILDANFIISGTSIG 1619  
 QY 1621 QFELICDKXKNIQPYFINFKIKETSITLYVGNRQNLIVEPSYHLDSDSGNISSTVINFSOK 1680  
 DB 1620 QFETICDENDNIQPYFIKFNLTETNTLYVGNRQNMIVPEPNYDLDSDSGNISSTVINFSOK 1679  
 QY 1681 YLYGIDRVYVNVKVIAPNLYT 1700  
 DB 1680 YLYGIDSCVKNVVISPNYIT 1699

RESULT 2

US-08-405-496A-10

; Sequence 10, Application US/08405496A

; Patent No. 5919665

; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.

; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

; TITLE OF INVENTION: NEUROTOXIN

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/405,496A  
 FILING DATE: 16-MAR-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 25-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/161,907  
 FILING DATE: 02-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: INGOLIA, DIANE E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPHD-01308  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2366 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-405-496A-10

Query Match 77.6%; Score 6733.5; DB 2; Length 2366;  
 Best Local Similarity 76.4%; Pred. No. 0;  
 Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

QY	1	MNLVKAQIQKVVYKFRQEDVEYVAILNALDEYHNMSESSVVEKYKLDKNNLTNYL	60
DB	1	MSLVNRKQLKXANVRFRQEDVEYVAILDALDEYHNMSESSVVEKYKLDKNNLTNYL	60
QY	61	NTYKSGRNKALKKPKFYLTMVELEKNNSLTPVEKNLHFHWIGQINDTAINYNQWKD	120
DB	61	DTYKSGRNKALKKPKFYLTMVELEKNNSLTPVEKNLHFHWIGQINDTAINYNQWKD	120
QY	121	VNSDYTVKPYDSNAPLINTLTKTIVESATNNTLSEFRENLDPEFDYKFKRMEIYY	180
DB	121	VNSDYVNVVYDSNAPLINTLTKTIVESAINTLSEFRENLDPEFDYKFKRMEIYY	180
QY	181	DKQKHFIDYKSOIEENPEFIDNIKTYSNEYSKDLEALKNYTEESLUNKITANNNDI	240
DB	181	DKQKFINYKQARENPELIDDIKTYLSNEYSKDELTYTEESLUNKITQNSGNDV	240
QY	241	RNLKPADBDLVRNLQNELVERWNLAAASDILRISMLKEDGGVLDVILPGIQDPLFKS	300
DB	241	RNFEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGMYLDVDMPLGQDPLFES	300
QY	301	INKPSITNTSWMKLEAIMKYKEYIPGTYTSKNFMDLDEEVORFESALSSKDKSEIF	360
DB	301	IEKPSVTVDFWEMTKLEAIMKYKEYIPETSEHFDMLDEEVQSSPFESVLASKDKSEIF	360
QY	361	LPDLDDIKVSPLEKAFAPANNVINQALISLKDYSYCSDLVINQIKNRYKITLNDNLAPSE	420
DB	361	SSLGMEASPLEKAFAPANNVINQALISLKDYSYCSDLVINQIKNRYKITLNDNLAPSE	420
QY	421	GTDFNTTMMKIFGDKLASINEDNMFMKIKITNYLVKGFAPDVRSTRINLSGGVYTCAYQD	480

DB	421	DNDFNTTNTTIDSIABANADNGRMMELGKYLVRGFFPDVKTNTINLSGPEAYAAAYQD	480
QY	481	LLMFKDNSTNHLLEPELRNPEFPKTKISQLEQEIITSLWSFNOARAKSQREYKKGYPE	540
DB	481	LLMFKEGWNHLEIADLRNPEISKTNISQTEQEWASLWSFDDARAKAQFBEYKKNYFE	540
QY	541	GALGEDDNLDPANQTVLDKDYVSKKILSMKTRNKEYIHYIVOLQDKISYBASCNLFK	600
DB	541	GSLGEDDNLDPANQTVLDKDYVSKKILSMKTRNKEYIHYIVOLQDKISYBACNLFK	600
QY	601	DYSSILYQKNTEGSETAYYYVADABEIKDYKIPYQISNKRNIKLTFIGHGSEFNT	660
DB	601	TPYDVLFPQNIJEDSEIAYYYPGDEIQEIDKYKIPSIISIRKIKLTFIGHGKDEFT	660
QY	661	DTFANLDVDSLSSEIETILNLAKADISPKYIEINILGCNMFYSYVABEYTPGKLLKIK	720
DB	661	DIFAGFDVDSLSSEIETILNLAKADISPKYIEINILGCNMFYSYVABEYTPGKLLKIK	720
QY	721	DRVSELMPSISQDSITVSANQYEVRAINBEGKREILDHSGKWINKESIIKDISKEYISF	780
DB	721	DKISELMPSISQDSIIVSANQYEVRAINBEGKREILDHSGKWINKESIIKDISKEYISF	780
QY	781	NPENKILVSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDQIIVER	840
DB	781	NPENKILVSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDQIIVER	840
QY	841	IEEAKNLTSDSINYIKNEPKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIP	900
DB	841	IEEAKNLTSDSINYIKNEPKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIP	900
QY	901	INKETGNSIFITEKEIFSEYATHISKEISNICKDIFONVNGKLVKKNLDAAEVNTLN	960
DB	901	INKETGESIFVETEKTFSEYANHITBELSKIGTIFDVTNGKLVKKNLDTTHEVNTLN	960
QY	961	SAFFIQSLIEYNTTKESLSNLSVAMKVQYIAQLFSTGLNTITDASKVELVSTALDETID	1020
DB	961	RAFFIQSLIEYNTTKESLSNLSVAMKVQYIAQLFSTGLNTITDASKVELVSTALDETID	1020
QY	1021	LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPLROEIAKIGIMAVNLTAASATVIT	1080
DB	1021	LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPLROEIAKIGIMAVNLTAATVIT	1080
QY	1081	SALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGFTLLD	1140
DB	1081	SSGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHISLVEGFTLLD	1140
QY	1141	DKIMPDQDLVLSIEDFNNSITLCKEIRWABGSGHTLTDIDHFPSSPSITVRKPL	1200
DB	1141	DKIMPDQDLVLSIEDFNNSITLCKEIRWABGSGHTLTDIDHFPSSPSITVRPHL	1200
QY	1201	STYDVLNTKKEIKDFSKDLMLPNAPNRFVGMGTGTCFRLDNDGTLLDRIDHVEG	1260
DB	1201	STYDVLVLEOKGELDUSKDLMLPNAPNRFVGMGTGTCFRLDNDGTLLDRIDHVEG	1260
QY	1261	QPYRWYAFIADALITLKPREDTNVRINLDGNTRSPVITTEQIRKNLSYFYGSG	1320
DB	1261	QPYRWYAFIADALITLKPREDTNVRINLDGNTRSPVITTEQIRKNLSYFYGSG	1320
QY	1321	GSYSLSLSPYNNIDNLVENDTWVIDVNVVKNITTESDEIQKELIENILSKLNIEN	1380
DB	1321	GYALSLSQYNNGINIELSESDVMIIDVNVVDRVTIESDKIKKGLIEGILSTLSIEN	1380
QY	1381	KILNHNHTINFGDINESNRPISLTSFLEIDINIIIEIDLVSYSKILLSGNCKLIENS	1440
DB	1381	KILNHNHTINFGDINESNRPISLTSFLEIDINIIIEIDLVSYSKILLSGNCKLIENS	1440
QY	1441	SDIOQKIDHIGFNGHQYIFYSYIDNETKXNGPDYDKKGLTAETESNEGIIRNIYMP	1500
DB	1441	NHIQKIDYIGNSLQKNIYFVSDEKENGFGNGSTKGLFVSELPDVLISKVYND	1500
QY	1501	DSNNLFYSSKDLKDIRINKGDKVLLIGNYFKDKMKSLSFTIEDTNTIKLVGVLDN	1560
DB	1501	DSKPSFGYVNNLKDKVITKONVNILTYLKKDDIKISLSLTLQDEKTIKLSNVHLS	1560



Db 961 RAFFIQSLIEVNSSKESLNSLVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETID 1020  
Qy 1021 LPTLSEGLPIITATIIDGVSIGAAIKELSETNDPLLRQIEAKIGIMAVNLTAARSTAVT 1080  
Db 1021 LPTLSEGLPIITATIIDGVSIGAAIKELSETSDPLLRQIEAKIGIMAVNLTAARSTAVT 1080  
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIOLQKATKVIDYFKHISLAETEGAFITLID 1140  
Db 1081 SSGIASGFSILLVPLAGISAGIPSLVNNELVLRQKATKVVDYFKHISLVETEGVFTLID 1140  
Qy 1141 DKIMPQDVLVSEIDFNNSITLQKCEIWRAGSGHTLTDIDHFPSSPSITVRKPNL 1200  
Db 1141 DKIMPQDVLVSEIDFNNSITLQKCEIWRAGSGHTVTDIDHFPSSPSITVRKPNL 1200  
Qy 1201 SYDVLNITKKEKIDFSKOLVLPNAPNRPVFGEMGTGFSRLDNDGKLLDIRDHVEG 1260  
Db 1201 SYDVEVQKEELDSKOLVLPNAPNRPVFAVETGTPGLRSLDNDGKLLDIRDHVEG 1260  
Qy 1261 QPYWRYFAFIADALITKLPYEDTNVRINLGNTRSFIVPITTEQIRKNLSYFYSGG 1320  
Db 1261 BPYWRYFAFIADALITKLPYEDTNVRINLGNTRSFIVPITTEYIREKLSYFYSGG 1320  
Qy 1321 GSYLSLSPYNNIDNLVENDTWIDVNVVKNITISDEIQKGLNLSKLNIEDN 1380  
Db 1321 GYALSLSQYNNGINELSESVDWIIDVNVVRDVTISDKIKKGLIEGILSTLSIBEN 1380  
Qy 1381 KIILNHTINFYGDINESRFTLSFLEIDNIIIEIDLVSYSKILLSGCMKLIENS 1440  
Db 1381 KIILNSHEINFGEVNGSGVSLTFSILEGINAIEVDLLSKSYKLLISGELKILMLNS 1440  
Qy 1441 SDIOQKIDHIGFNGHQKIPFYSYIDNETKNGFDYDKKEGLTAETFSNESIIRNIYMP 1500  
Db 1441 NHIQKIDYIGNSLQKIPFYSVDSEKENGFGSTKEGLFVSELDPVVLISKVYMD 1500  
Qy 1501 DSNLPYSSKDLKDIRINKGDKVLLIGNYPKDMKVSLSFTIEDTNTIKLNGVYLDEN 1560  
Db 1501 DSKPSFGYSSNNLKDKVITKDNVILTGYYLKDDIKISLSLTLODEKTIKLSVHLDES 1560  
Qy 1561 GVAQILKFNNAKSALNTSILNPLFLESINIKIFVNLDPNIEFTLDTNFTISGNSIG 1620  
Db 1561 GVAELKFNMR-KGNTNTSDLSLMSLESNNIKSIFVNLQSNIKFILANFTISGTSIG 1619  
Qy 1621 QPELICDKDKNTQPYFINKETSYLVGNRQNLVPEPSYHLDDSGNISSTVINFSOK 1680  
Db 1620 QEFICDENDNTQPYFINKETSYLVGNRQNLVPEPSYHLDDSGNISSTVINFSOK 1679  
Qy 1681 YLYGIDRYNVKVIAPNLYT 1700  
Db 1680 YLYGIDSCVNVKVISPNLYT 1699

RESULT 4  
US-08-957-310-10  
; Sequence 10, Application US/08957310  
; Patent No. 6365158  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Kink, John A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; TITLE OF INVENTION: DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Medien & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,310  
; FILING DATE: 23-OCT-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-957-310-10  
  
Query Match 77.6%; Score 6733.5; DB 4; Length 2366;  
Best Local Similarity 76.4%; Pred. No. 0;  
Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;  
  
Qy 1 MNLVNAQLQKVVYKFRIOEDYVAILNALSEYHNMSSESVKYLKLDIINNLTADNYL 60  
Db 1 MSLVNRKQLEKMANVRPTQDEYVAILDALSEYHNMSSENTVVEKYLKLDIINLTDIYI 60  
  
Qy 61 NTYKSGRKNALKKFKEYLTWVELELKNLSITPVEKNLHFVIGQINDTAINYNQWKD 120  
Db 61 DTYKSGRKNALKKFKEYLTWVELELKNLSITPVEKNLHFVIGQINDTAINYNQWKD 120  
  
Qy 121 VNSDYTVKFVDSNAFLINTLKKTIVESATNTNLTLESPRENLNDPFDYKFKRMEIYY 180  
Db 121 VNSDYTVKFVDSNAFLINTLKKTIVESATNTNLTLESPRENLNDPFDYKFKRMEIYY 180  
  
Qy 181 DKQHFIDYKSOIEENPEFIIDNIITKYLSEYKOLEALNKYIEESLNKITANNGNDI 240  
Db 181 DKQNFINYKAQREENPELIIDDIKTYLSNEYSKIDEIKNYIEESLNKITANNGNDV 240  
  
Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDLIRISMLKEDGGVYLDVILPGIQDLPKS 300  
Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDLIRISMLKEDGGVYLDVILPGIQDLPKS 300  
  
Qy 301 INKPSITNTSWEMIKLEAIKMYKEYIPGYTSKPNFDMDEEVQSFESALSKSKSEIF 360  
Db 301 IEKPSVTVDFWMTKLEAIKMYKEYIPGYTSKPNFDMDEEVQSFESALSKSKSEIF 360  
  
Qy 361 LPDLDIKVSPLEKIAFANNVINOALISLKDYSCLVINOIKNRYKILNDNLNPSINE 420  
Db 361 SSGDMEASPLEKIAFNKGIINGQLISVKDYSCLVINOIKNRYKILNDNLNPSINE 420  
  
Qy 421 GTDFNTMTKIFSDKLASISNEDNMFMKITNYLKVGFADPVRSTINISGPGVYTGAVQD 480  
Db 421 DNDFNNTTNTTIDSIDMAEADNCRFMELGKLVGFFPDVKTINISGPEAYAAAYQD 480  
  
Qy 481 LLMFKDNTNHLLEPELNRNEFPKTKISQLEQITSLMSFNQARAKSOEEYKGVPE 540  
Db 481 LLMFKEGSMNHLLEADLRNFEISKTNISQLEQITSLMSFNQARAKSOEEYKGVPE 540  
  
Qy 541 GALGEDDNLDPQNTVLDKDYVSKILLSMKTNRKEYIHYIVQLQGDKISYEASCNLFPSK 600  
Db 541 GALGEDDNLDPQNTVLDKDYVSKILLSMKTNRKEYIHYIVQLQGDKISYEASCNLFPSK 600

Db 541 GSLGEDNDLFSQNIIVVDKEYLLEKISSIARSSERYIHYIVLOQDKISYEACNLFAK 600  
Qy 601 DVSSTLYOKNTEGSTAYYYVADAEIKEDIKYRIPYOISNKRNIKLTFIGHGSEFNT 660  
Db 601 TPYDSVLFQNIEDSIAIYVNGDGEIDKIKIPSIIDRPKIKLTFIGHGDEFT 660  
Qy 661 DTFANLDVDSLSSEIFTLINLAKADISPKYIIBINLLGCNMFYSIYABETYPGKLLKIK 720  
Db 661 DIFAGFDVDSLSTEIEAADLAKEDISPKSIEINLLGCNMFYSINVEETYPGKLLKVK 720  
Qy 721 DRVSELMPSISQDSITVSANQVEVRINEEGKEIIDLHSGKWNKESIIKDISKEYISF 780  
Db 721 DKISELMPSISQDSITVSANQVEVRINSEGRRELLDHSGEINKESIIKDISKEYISF 780  
Qy 781 NPEKNKIIKSVKYLHSLTLLQEIIRNANSDDIDLEKKVMLTECEINVASNIDROIVEGR 840  
Db 781 NPEKNKITVKNLPELSLTLQEIIRNANSDDIELEKVMLTECEINVISNIDTQVEER 840  
Qy 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900  
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDALCDLKQNELEDHFSFEDISEDEGFSIRF 900  
Qy 901 INKETGNSIFITEKEIFSEYATHISKEISNTKOTIFDQVNGKLVKVNLDAAHEVNTLN 960  
Db 901 INKETGESIFVETEKTFSEYANHITEISKIKGTIFDTVNGKLVKVNLDTHEVNTLN 960  
Qy 961 SAFFQSLIENYTKESLNSVAMKVQVYAOIFSTGLNTITDASKVVELVSTALDETD 1020  
Db 961 AAFQSLIENYTKESLNSVAMKVQVYAOIFSTGLNTITDAAKVELVSTALDETD 1020  
Qy 1021 LAPTISEGLPIATIIDGVSGLAAIKELSETNDPLLRQIEAKIGIMAVNLTAASATV 1080  
Db 1021 LAPTISEGLPIATIIDGVSGLAAIKELSETNDPLLRQIEAKIGIMAVNLTAATIT 1080  
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGFTLLD 1140  
Db 1081 SSIAGSFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHVSIVETEGVFTLLD 1140  
Qy 1141 DKILMPQDLVISEIDFNNSITLGKCEIWRAGSGHGLTDDIDHFFSPSITVYKPKWL 1200  
Db 1141 DKILMPQDLVISEIDFNNSIVLGKCEIWRAGSGHGLTDDIDHFFSAPSITVREPHL 1200  
Qy 1201 SIYDVNLNKKKIDFSDKMLVLPNAPNRFVGMGTGFRSLDNDGTLKLLRDRDHYEG 1260  
Db 1201 SIYDVLEVQKEELDLKDLMLVLPNAPNRFVAFETGTPGLSLENDGTLKLLRDRDHYEG 1260  
Qy 1261 QFYWRYFAFIADALITFLKPRVEDTNVRINLGNTRSFIVPVTIETQIRKNLSYFYSGG 1320  
Db 1261 EFWRYFAFIADALITFLKPRVEDTNIRINLGNTRSFIVPVTIETIREKLSYFYSGG 1320  
Qy 1321 GSYLSLSLSPNNMIDNLVENDTWTVDVNVKNTIESDEIOKGELIENILSKLNIEN 1380  
Db 1321 GTYALSLSQYNNMGINTELSSEVDWIIIDVNVVDRDVTIESDKIKGDLIEGILSTLSIEN 1380  
Qy 1381 KIILNHTINFYGDINESNRFSITLESILEDINIIEIDLVSQYKILLSGCMKLIENS 1440  
Db 1381 KIILNSHEINFSEVNGSFGVSLTSLIEGINAIEVLLSKYKILLSGELKILMUNS 1440  
Qy 1441 SDIQKIDHIGFNGEHQKIFYSYIDNETKYNFIDYSKKEGLFTAEFNSIESIIRNIYMP 1500  
Db 1441 NHIQKIDHIGFNGESELQKNIPIYSFVSEKENGFGINGSTKEGLFVSELDDVLLISKVMD 1500  
Qy 1501 DSNLFIYSKDLDIRIINKGDKVLLIGNYFKDDMKVSLSTIEDTNTIKLNGVVLDEN 1560  
Db 1501 DSKPSFGYSSNNLKVVKVITKONVNLTGYLLKDDIKISLSLTLQDEKTIKUNSVHLDGS 1560  
Qy 1561 GVAQILKFWNNAKSALNTSNLNFLESINIKNIFYNLDPNIEFTLDTNFTISGNSIG 1620  
Db 1561 GVAEILKFWNR-KGNTNTSDSLMSFLESNNISIFVNFLOSNIKFLIDANFIISGTSIG 1619  
Qy 1621 QFELICDKDKNIQPIYFNFKIETSYTLVYGNRQNLIVPSPVHLDGNSISSTVINFSQK 1680  
Db 1620 QFEFICDENDNTQPIYFNFKIETSYTLVYGNRQNMIVPNVDLSDGSDISSTVINFSQK 1679

Qy 1681 YLYGIDRYVNVKVIAPNLYT 1700  
Db 1680 YLYGIDSCVNVKVISPNLYT 1699  
RESULT 5  
US-10-011-366-10  
; Sequence 10, Application US/10011366  
; Patent No. 6573003  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; Kink, John A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/011.366  
; FILING DATE: 16-NO. 6573003-2001  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957.310  
; FILING DATE: 23-OCT-1997  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-011-366-10  
Query Match 77.6%; Score 6733.5; DB 4; Length 2366;  
Best Local Similarity 76.4%; Pred. No. 0;  
Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

Qy 1 MNLVNAQLQKVVYKFIQDEYVAIINLAEYHNMSESSVVEKYLKDKINNLTDNYL 60  
Db 1 MSLVNRKQLEKMANVFRTOQEYVAIILDALEEHNMSENTVVEKYLKDKINSLTDIYI 60  
Qy 61 NTKYKSGRNKALKKFEKYLTMVELELKNSTLPVEKNLHFHWIGGQINDTAINYNQWKD 120  
Db 61 DTKYKSGRNKALKKFEKYLTMVELELKNSTLPVEKNLHFHWIGGQINDTAINYNQWKD 120  
Qy 121 VNSDYTVKRVYDSNAFLINTLTKTIVESATNTLTESFRENLDNDFDYNKFKYRKRMEIY 180



[illegible]

RESULT 6  
US-08-480-604A-6  
; Sequence 6, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994



US-08-405-496A-6  
; Sequence 6, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2710 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-405-496A-6

Query Match 51.4%; Score 4458; DB 2; Length 2710;  
Best Local Similarity 50.7%; Pred. No. 1.5e-261;  
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;  
QY 1 MNLVNAQOLQKVVYKFRQDEYVAILNALREYNMSSVVEKYLKLDINLTDNYL 60  
DB 1 MSLISKEELIKUAY-SIRPRENEYTLTNDLDEYNKLTNNENKYLQKUNESIDVPM 59  
QY 61 NTYKSGRNKALKCFKXYLTMEVLELKNNSLTPVEKNLHFHWIGGOINDTAINYNQWKD 120  
DB 60 NKYKTSRRNALSNLKDKLKEVILKNSNTSPVEKNLHFHWIGGEVSDIALEYIKQWAD 119  
QY 121 VNSDYTVKVPYDSNAPLINTLKTIVESATNTNLTSLFRENLDNPDYNNKFRKRMELIY 180  
DB 120 INAEYNIKLYWSEAFVNTLTKAIVESSTTEALQLEEEIQNPDPNNKFKYKGRMEFTY 179  
QY 181 DKQKHFIDYKSOIENPEFIIDNIKTVLSNEYSKDLKALNKLYEESLKNKITANNNDI 240  
DB 180 DRQKRPINYSQINRPVPTTDDIILKSHLVSEYNRDETVELSEYRTNSLRKINSNHGDI 239  
QY 241 RNLEKFAEDLRLVNLQELVERWNLAAASDIIRISMLKEDGGVYLDVILPGIQPDLFKS 300

DB 240 RANSLFTEQELLNTYSQELLNRGNLAAASDIIVRLALKNPGGVYLDVMDLFCIHSDFLTKT 299  
QY 301 INKPDSTINTSWEMIKLEAIWKYKEYIPCYTSKNPMDLDEEVQSFESALSKSKSKSEIF 360  
DB 300 ISRFSGICLDREMIKLEAIWKYKYNNTYSENFDKLDQOKNFKLIIESKSEKSEIF 359  
QY 361 LPDDIKVSPLEVKIAFANNVINQALISLKDSCYSDIIVINOIKRYKILNDNLNPSINE 420  
DB 360 SKLENLNVSDLEIKIAPALGSVINQALISKQSVYLTNLVIBQVKRYQFLNQHLPAYES 419  
QY 421 GTDFTNTMKIPSDKLASISNEDNMFMKIITNYLKVGFAPDVRSTINISGCVYTGAYQD 480  
DB 420 DNNFTDTTKIFHDSLFNSATAENSMTLKIPYQVGFMPPEARSTISLUSGAYASAYD 479  
QY 481 LLMPKDNSTNTHLLEPELURNEFPKTKISQLTQEBITSLWGFNQAARAKSQPEYKKGYPE 540  
DB 480 FINQENTIEKTLKASDLIEKFPENNLSQLTQEINSLWSPDQASAKYQEKYVRYDTYG 539  
QY 541 GALGEDDNLDPQNTVLDKQY-VSKKILSS--MTRNKYIHYIVQLQGDKISYEASCNL 597  
DB 540 GSLSEDNGVDFNKNTALDKNYLLNKNIPSNVVEAGSKYVHYIILQCGDDISYEATCNL 599  
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEDKVRIPYQISNKNIKLTFIGHCK 655  
DB 600 FSKNPKNSIIQRNM--NESAKSYPLSDGSEIIELEKNYRIPERLKNKEKVKVTFIGHGK 657  
QY 656 SEENTDTFANLDVDSLSEIETILNLAKADISPKYIIEINLGCNMFYSIYAEETYPGKL 715  
DB 658 DEFNTSEFARLSDVSLSEISFDTIKDLSIPKNVEVNLGCNMFSDFNVEETYPGKL 717  
QY 716 LLKIKDRVSELMPGISQDSITVSANQYEVRIINEEGKREILDHSGKWKINKEESIIDKSSK 775  
DB 718 LLSIMDKITSTLPDVNKNSTIGANQYEVRIINSEGRKELLAHSGKWKINKEEAIMSDLSK 777  
QY 776 EYISFPNPKENKIIVKSKYLHBLSTLQLOIRNANNSOIDLKSKVMLTCEINVAENIRQ 835  
DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFTLNNLKLNISSIGDY 837  
QY 836 IVEGRIBEAKNLTSDSINYIKNEPKLIESISDSLYDLKHQGLDDSHFISPEIDIKTENG 895  
DB 838 IYEEKLEPVKNIIHNSIDDLDEFNLEENVSDLEYELKLNLDKEYLISFEDIESKNST 897  
QY 896 FRIRPINKETGNSIFETEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKKNVLDAAHE 955  
DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQLDHTSQ 957  
QY 956 VNTLNSAPFIQSLIENYNTTKESLSNLSVAMKVQVYLAQLPSTGLNTITDASKVVELVSTAL 1015  
DB 958 VNTLNAAFFIQSLIDYSSNKDVLNLDLSTSVKQVLAQLFSTGLNTIYDSIQVNLNISNAV 1017  
QY 1016 DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPDLRQETAKIGIMAVNLTAAS 1075  
DB 1018 NDTINVLPTTIEGPIVSTIILDGINIGAAIKELDEHDPDLKKELEAKGVGLAINNLSUI 1077  
QY 1076 TAITVSALGIASGFSILLVPLAGISAGIPSLVNNELIQLDKATKVIDYFKHISLAETEGA 1135  
DB 1078 AATVASIVGIGAEVTFILPLAGISAGIPSLVNNELILHDKATSVVNFNHLSESKYCP 1137  
QY 1136 FTLLDDKIIMPDDLVLSIEDFNNNISITLKGCEIWRAGGSGHTLTDDIDHFFSPSPITY 1195  
DB 1138 LKTEDDKLVPDLDLVISEIDFNNNISIKLGTGNILAMEGGSGHTVTGNIDHFFSPSPIS 1197  
QY 1196 RKPWLSIYDVLNLIKKEKIDFSKOLMVLNAPNRFGEYMGWTPGFRSLDNDCGTKLLDR 1255  
DB 1198 HIPSLSIYSALGIEFENLDFSKKIMMLPNAPSRVFWWETGAVPGLSRENDGTCLLDSIR 1257  
QY 1256 DHYEQFYWRYPFAFTADALITKLKRYEDTNNRINLDGNTRSFIVPVITTEQIRKNLSYS 1315  
DB 1258 DLYPKFYWRYPFAF-DYAITTLKVPYEDTNIKIKLDKDRNFIMPTITTNIRNKLVSYS 1316  
QY 1316 FVSGGYSLSLSPPNMDNLNVDNVDVKNITTESDEIQGELIENILSKL 1375  
DB 1317 PDGAGGTYSLLSSYPISNTINLSKDDLWIFNIDNEVREISITENGTIKKGKLIKDVLSKI 1376

QY 1376 NIEDNKKIILNHTINPYGDINESRFLSTFSTILEDINIIIEIDLVSYSKILLSCNWK 1435  
Db 1377 DINKNLIIGNQDTISGDINDKDRIFLUTCEDDKISLIIIEINLVAKSYLLSGDKNY 1436  
QY 1436 LIENSSDIOQKIDHIFNGEHOKYIFYSYID-NETKINGFIDYSKKEGLFTAESNESII 1494  
Db 1437 LISNLSTIEKINTGLD--SKNIAVYVDESNNKYFCAL-----SKTSQKSII 1483  
QY 1495 RNIYMPDSNL-----FIYSSKDL--KDIRINKGVDKLLIGNYFKD---DMKVSLS 1541  
Db 1484 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVFWMKDDINTIGKYVVDNNTDKSIDFS 1541  
QY 1542 FRIEDTNTIKLGVLDENGVAQILKFMNNAKSALNTSNLMELESINIKNIFNNLDP 1601  
Db 1542 ISLVSKNQVKNGLYVNSYSLDFVKNOSGHNHTSNFMNLFNDISFWKLGFPE--- 1598  
QY 1602 NIEFILDNTFIISGNSIQGFELICDKOKNIOPYFINFKIKETSVTLVYGNRQNLIVBPS 1661  
Db 1599 NINFVIDKYFTLVGKTNLGVVEFICDNNKNIIDYFGWKTSKSTIFSGNGRNVVVEPI 1658  
QY 1662 YHLDSSGNSSTVINPQSKYLGVIDRYVKNVIAPNLYT 1700  
Db 1659 YN-PDTGEDISTSLDPSYEPFLYIDRYINKVLIAPDLYT 1696

## RESULT 8

US-08-915-136-6  
Sequence 6, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-915-136-6

Query Match 51.4%; Score 4458; DB 3; Length 2710;  
Best Local Similarity 50.7%; Pred. No. 1.5e-261;  
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY 1 MNLVNAQLOKVVYKFRIOEYVAIILNAAEEYHNHSESSVVEKYLKLDINNLTDNYL 60  
Db 1 MSLISKEELIKLAY-SIRPRENEYKITLTDNDEYKLTNNNNENKYLQKKLINESIDVFM 59  
QY 61 NTYKSGRNKALKKPKKYLTMVLELKNNSLTPVEKNLHFTWIGQINDTAINYNQWKD 120  
Db 60 NKYTSSNRNALSNLKDKILKEVILIKNSNTPVEKNLHFTWIGGEVSDIALEYIKQWAD 119  
QY 121 VNSDYTVKVVDSNAFLINTLTKTIVESATNTNLTESFRENLDNPDFFYKPKRMEIY 180  
Db 120 INAEYNIKLYDSEAFVLTLLKAIIVESSTTEALQLLEEIQNPQFDMNKFKKRMFIY 179  
QY 181 DKQKHFDYKSOJTEENPEFIIDNIIKTYLSNEYSKDLKALNKYIEESINKTANGNDI 240  
Db 180 DRQKRFINYKQINKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGDI 239  
QY 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGCVYLDVLDVILPQIDPLFS 300  
Db 240 RANSLFTEQELLNIYSQELLNARGLNLAASDIVRLLAKNFGVYLDVMDLFGIHSDLPT 299  
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKPMDEEVSQSFESALSKSKSEIF 360  
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQDLKONKFLIIESKSEIF 359  
QY 361 LPDDIKVSPLEVIAPANNVINOALISLSDSVCLAVINQIKRYKILNDNLPSINE 420  
Db 360 SKLENLVNLDLEIKIAFGVINOALISKOSYLTNLVIEQVKRYQLNQHLPALIES 419  
QY 421 GTDFNTTMKIFSDKLASINEDNMFMIKITNYLKVGFPADVPVRSTINLSGPGVYTCAYQD 480  
Db 420 DNNFTDTTKIFHDSLFNSATAENSMTLKIAPYLQVGFMPPEARSTISLSGPGAYASYD 479  
QY 481 LLMFKNSTNTHLLEPELRNFEPPKTKISQTEQBITSLMSFNQARAKSQEYKGYFE 540  
Db 480 FINLQNTIEKTLKASDLIEFKFPENNLSQLTEQIEINSLWSFDQASAKYQEKYVRDVTG 539  
QY 541 GALGEDNDLFAONTVLDKDY-VSKKILSS--MKTRNKEYIHYIYVLOGDKISYASCNL 597  
Db 540 GSLSEDNGVDNKNYNTALDKNYLLNNKIPSNVVEAGSKNYHYIYIQLQGDIDISYATCNL 599  
QY 598 FSKDPYSSILYQKNIEGSETAYVYVAD--AEIKEDKRYIPYQISNRKNLTLTIGHGK 655  
Db 600 FSKPKNSIIIIQRNM--NESAKSYFLSDGDSIELELNKRYPERLKNKKEKVTYIGHGK 657  
QY 656 SEFNFTDTFANLDVDSLSSEIETILNLAkadISPKYIEINLGCNMFYSYIYAEETYPGL 715  
Db 658 DEFNTSEFARLSVDSLSNEISSFLDTIKDISPKNVEVNLGCGNMFYSYDFNVEETYPGL 717  
QY 716 LLKIKDRVSELMPSISQDSITVSANOYEVNRNEKREILDHSGKWNKESIIIDKSSK 775  
Db 718 LLSIMDKITSTLPDVNKNISITIGANQYEVNRNEKREILDHSGKWNKESIIIDKSSK 777  
QY 776 EYISFPNKENKIIIVKSKYLHELSTLLOIBRNANSDDIDLEKVKVMLTCEINVASNIDRQ 835

778 EVIFFSDINUKLAKSKNIPGLASISEDIKTLLDASVSPDTKFLINLNKLIAMIESIGDY 837  
QY  
836 IVEGRIEEAKNTSDSINTYKNEFKLIESISLYDLKHONGLDSDHFSFEDISKTEG 895  
Db  
838 IYKLEPVNIIHNSIDDLIDFNLENVSBELYELKKNLDELKILSFEDISKNT 897  
QY  
896 FRIRPINKETGSIPIETEKEIFSEYATHISKEISNIKDTIFDNVNGKLKVKVNDAAHE 955  
Db  
898 YSVRFINKSGESVYVETEKEIFSKYSEHIKTEISTIKNSIITDVNGLNLDNIQLDHTSQ 957  
QY  
956 VNTLASAFIQLIENYTTKESLSNLSVAMKVQVQAOLFSTGLNTITDASKVVELVSTAL 1015  
Db  
958 VNTLNAAFIQLIDVSSKNKDLNLDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLTNAV 1017  
QY  
1016 DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQETAEKIGIMAVNLTAAS 1075  
Db  
1018 NOTINVLPITIGIPVITILGINLGAAIKELLDHDPKKELEBAKVGVIAINWSLSI 1077  
QY  
1076 TAVTSALGIAGFSILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGA 1135  
Db  
1078 AATVASIVGIGAETIFLPIAGISAGIPSLVNNELIHDKATSVVYFNHLSSEKKGYP 1137  
QY  
1136 FTLLDDKIIMPDDLVLSIDFNNSITLCKEIPRAEGSGHTLTDDIDHFFSPSIY 1195  
Db  
1138 LKTEDDKILVPIDDLVISSIDFNNSIKLGTNCILAMEGSGHTVTGNIDHFFSPSIS 1197  
QY  
1196 RXPWLSIYDVNLKKEKIDFSDKLMVLPNAPRVFGYEMGWTPGPRSLNDGTKLLDRIR 1255  
Db  
1198 HPSLSIYSAIGIETENLDFSKIMLPNAPRVFWETGAVPGLSLNDGTRLLDSIR 1257  
QY  
1256 DHYEQFYWRYPFAITADALITKLKPYEDTVNRINLDGNTRSFIIVPITTEQIRKNLSYS 1315  
Db  
1258 DLYPGKFYWRYPFA--DYAITTLKPVEDTNIKIKLDKOTRNFIMPTITITNEIRNKLVSYS 1316  
QY  
1316 FYGSGSYLSLSPYNNIDNLVENDTVIDVNVKNITIESBIOQGLIENILSKL 1375  
Db  
1317 PDGAGTYSLLSSYPISITNINLSKDDLWIFNIDNEVREISIEGNTIKKGLIKOVLSKI 1376  
QY  
1376 NIEDNKIILNNHTINPYGDINESNRPISLTSFISLEDINIIIEIDLVSYSKYLSCNCK 1435  
Db  
1377 DINKKLIIGNOTIDFSGDIDKDRYIFLTCDDKISLIIELNLVAKSYLSLSDGRNY 1436  
QY  
1436 LIENSDDIOQKIDHGFNGEHOKYIPYSYID--NETKINGFIDYSKKEGLTAEFSGNESII 1494  
Db  
1437 LISLNSLTIEKINTGLD---SKNIAYNTDESNNKYFCAL-----SKTSQKSI 1483  
QY  
1495 RNIIYDPSNNL-----FYSSKDL--KDIRINKGVKLLIGNYFKD---DMKVSUS 1541  
Db  
1484 H--YKOSKNILEFYNDSTLEFNSKDPIAEDINVFMKDDINTITGKYVDNNTDKSIDPS 1541  
QY  
1542 FTIEDTNTIKLGVYLDENGVAQILKFMNNAKSALNTSNLNFLESINIKNIFYNNDLP 1601  
Db  
1542 ISLVSKNQVKNGLYNESVYSYLDVFNKSGHNTSNFNLFDLNTSFWKLFQFE--- 1598  
QY  
1602 NIEFLDNTFIISGNSIQOPELICDKOKNIOPYFINFKIKETSITYLVGNRQNLIVEPS 1661  
Db  
1599 NINFVIDKYFTLVGKTNLGYVEFICDNNKNIIDYFGEMKTSKSTIFSGNGRNVVVEPI 1658  
QY  
1662 YHLDUSGNISSTVINPFSQKLVGIDRYVKNVJIAPNLYT 1700  
Db  
1659 YN--PDTGEDISTLSDFSEPLGYIDRYINKVLIAPDLYT 1696

RESULT 9

US-08-957-310-6  
; Sequence 6, Application US/08957310  
; Patent No. 6365158  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Kink, John A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPE  
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; TITLE OF INVENTION: DISEASE

NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FILING DATE: 23-OCT-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-957-310-6  
Query Match 51.4%; Score 4458; DB 4; Length 2710;  
Best Local Similarity 50.7%; Pred No.1.5e-261;  
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;  
QY 1 MNLVNAQLQKRVYKPRIQEDVYVAILNALBEEYHNMSESVYKYLKDKOINLTNYL 60  
Db 1 MSLISKEELIKLAY-SIRPRENEYKTLTLDYENKLTNNENKYLQKKNESIDVPM 59  
QY 61 NTKYSGRNALKKFEKVELTMEVLEKKNLTPVEKNLHFVIGQINDTAINYNQWD 120  
Db 60 NKYTSSRRNALSNUKDKILKEVILIKNSNTPVEKNLHFVWIGVEVSDIALEYIKQWAD 119  
QY 121 VNSDYTVFVYDSNAFLNTLTKTIVESATNTNLESFRENLDNDFDYKNKFKRMEIY 180  
Db 120 INAEYNIKLVWDSEAFVNTLKAIVESSTTEALQLLEEIQNPQFDNWKFKKMEFY 179  
QY 181 DKQGHFIDYKSQIBENPEFIIDNIKTYSNKSDEALNKYIEESLANKITANNNDI 240  
Db 180 DRQKRFINYKSQLNKPTVPTIDDIKSHLVSEYNRDETVELSYRTNSLRKINSNHGDI 239  
QY 241 RNLEKFADELVRNLQELVERNWLAAASDILRISMLKEDGGVYLDVILPGIQDLPKS 300  
Db 240 RANSLFTEQELNLYSQELNLRGNLAAASDILRLLALKNFGVYLDVILPGIHSDLFKT 299  
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPCYTSKNFDMLEDEEVQSFESALSSEKSEIF 360  
Db 300 ISRPSSIGLDREWEMIKLEAIMKYKEYINNYTSNFDFKLDQDQKQKFLIIBSEKSEIF 359  
QY 361 LPDLDIKVPSLEVKIAFANNVYNQALISLKDSCYSDLVINQIKRYKILNDNLNPSINE 420

Db 360 SKLENLVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLPALPES 419  
 Qy 421 GTDFTNTMKIPSKLASISNEDNMEMIKITNYLKUGFAPDVRSTINLSGRGVYTCAYOD 480  
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMEPEARSTISJSGPGAYASAYD 479  
 Qy 481 LLMFKDNSTNIHLLEPELANFPFKTKISQLEQETITSLWSFNQARAKQFEEYKGYPE 540  
 Db 480 FINLQENTIEKTLKASDLEFPENNLSQLTEQELNSLWSDQASAKYQFERYDYTG 539  
 Qy 541 GALGEDNDLFAQNTVLKDY-VSKILSS--MKTRNKYIHYIYVQLQDKISYEASCNL 597  
 Db 540 GSLSEDNGVDNFKNTALDKNYLLANKIPSNVVEEAGSKYVHYIIQLQGDIDISYEATCNL 599  
 Qy 598 FSKDPSYSSILYQKNVIGSETAYYYVAD--AEIKETDKYRIPYQSNKKNIKLFTFIGHCK 655  
 Db 600 FSKNPKNSIIIOIRM--NESAKSYFLSDGSELELNLKRYRIPERLKNKEKVKVYTFIGHCK 657  
 Qy 656 SEPTDTFRANLDVDSLSSSEIETILNLAKADISPKYTEINLLGNMFSYSIYAEETYPGKL 715  
 Db 658 DEFTSEFARLSVDSLNSISSFLDTIKLDISPANVEVNLGNMFSYDFNVEETYPGKL 717  
 Qy 716 LLIKDRVSELMPISQDSITSANQYEVRIINEEGREILDHSGKWINKEESIIDKISSK 775  
 Db 718 LLSIMDKITSTLPDVNKNSTIGGAGQYEVRIINSEGRKELLAHSGKWINKEEAIMSDLSK 777  
 Qy 776 EYISFNPKENKIIVKSKYLHELSTLQETRNANSSDIDLEKKVMLTECEINVASNIDRQ 835  
 Db 778 EYIFDSIDNKLKAKSKNIPGLASISEDITKLLDASVSPDTKFIILNNLKINIESIGDY 837  
 Qy 836 IVEGRIEEAKNTSDSINTYKNEFKLIESISLKYDLKHQGLDSDHSPISFEDISKTEG 895  
 Db 838 IYXKLEPVKNIIHNSIDIDLEFNLENVSDLEYELKLNLDKYLISFEDISKNST 897  
 Qy 896 FRIRFINKETGNSIFIETEKEIPSEYATHISKEISNIKOTIFDNVNGKLVKVNLDAAHE 955  
 Db 898 YSVRFINKNGESVYVETEKEIPSKYSEHITKEISTIKNSIITDVNGLLDNIQLDHTSQ 957  
 Qy 956 VNTLNSAFPIQSILIENTYKESLSNLSVAMKVQVYAQLSTGLNTITDASKVELYSTAL 1015  
 Db 958 VNTLNAAFFIQSLIDYSNNKVDNLSTSVKQVLAQLFSTGLNTIYDQIQLVNLISNAV 1017  
 Qy 1016 DETIDLLPTLSEGLPIATIIDQVSLGAAIKELSETNDPLLROEIBAKIGIMAVNLTAAS 1075  
 Db 1018 NDIINVLPITTEGPIVSTILDGINLQAAIKELDEHDPLLKELBAKVGVIAINMSLSI 1077  
 Qy 1076 TAVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGA 1135  
 Db 1078 AATVASIVGIGAEVTFILPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSSEKKYGP 1137  
 Qy 1136 FTLLDDKIIMPQDVLVSEIDFNNSITILGKEIWRABGSGHTLTDDIDHFPSSPSITY 1195  
 Db 1138 LKTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGSGHTVTGNDHFPSSPSISS 1197  
 Qy 1196 RKPWLSYDVLNIIKKEIKDPSKDLMLVPNAPRVFVEMGWTGPFSLNDGDKLLDRJR 1255  
 Db 1198 HIPSLSIYSAIGIETENLDPSSKIMMLPNAPRVFWETGAVPGLSLENDGTGLLDSIR 1257  
 Qy 1256 DRYEGOFYRYPAFIADALITKLKPRYEDTNVRINLDGNTRSPFVITTEQIRKNLSYS 1315  
 Db 1258 DLYPGKGYRWFYAFF-DYAITTLKPVYEDTNIKIKLDKOTRNFIMPTITITNEIRNKLVS 1316  
 Qy 1316 FYCGSGSYSLSPYNNMIDLNLVENDTWIDVNVVKITITESDIOKGLIENTLSKL 1375  
 Db 1317 FDGAGTYSLLSSYPISTNINISKODLMIFNIDNVEIREISENGITKKGKLIKOVLSKI 1376  
 Qy 1376 NIEDNKIILNNHINPYGDNESNRFTSITFLESDINIIIEIDLVSYSKYLISGNCMK 1435  
 Db 1377 DINKKLIIGNQITDSGDINDKORYFLTCLEDDKLSLIIELNLVAKSYLLSGLSKNY 1436  
 Qy 1436 LIENSDDIOQIDHIFGNGHQKIYFYSYID-NETKYNPIDYSKKEGLFTABFNSNESII 1494  
 Db 1437 LISLNSLTIEKINTGLID---SKNIAYNYTDESNNKYFGAI-----SKTSQSKSII 1483

Qy 1495 RNIVMPDNNL-----FTYSSKDL--KDIRIINKGDKVLLIGNYFKD---DMKVSLS 1541  
 Db 1484 H--YKOSKNILEFVNDSTLEFNSKDFTAEDINVPKDDINTYITCKYVDNNTDKSIDFS 1541  
 Qy 1542 FTIEDTNTIKLVGVDENGVAQILKFMNNAKSALNTSNLMNFLESINIKNIFYNLDP 1601  
 Db 1542 ISLVSKNQVKNGLYLNESVSYLDYFKVNSDGHHTSNFMNLFLDNISFWKLFGFE--- 1598  
 Qy 1602 NIEFLDTNFIISGNSIGQFELICDKQKQIOPYFINFKIKETSITLYVGNRQNLIVEPS 1661  
 Db 1599 NINFIIDYKFTLVGKTNLGYVEFICDNNKNIDIFYGEWKTSKSTISFGNGRNVVEPI 1658  
 Qy 1662 YHLDGSGNISSTVINFSOKYLYGIDRVYVKNVILAPNLT 1700  
 Db 1659 YN-PDTGEDISTSLDYSYEPYGLIDRYINKVLIAPDLTY 1696

RESULT 10  
 US-10-011-366-6  
 ; Sequence 6, Application US/10011366  
 ; Patent No. 6573003  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williams, James A.  
 ; Kink, John A.  
 ; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
 ; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
 ; DISEASE  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Medlen & Carroll  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/011,366  
 ; FILING DATE: 16-NO. 6573003-2001  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/957,310  
 ; FILING DATE: 23-OCT-1997  
 ; APPLICATION NUMBER: US 08/329,154  
 ; FILING DATE: 24-OCT-1994  
 ; APPLICATION NUMBER: US 08/161,907  
 ; FILING DATE: 02-DEC-1993  
 ; APPLICATION NUMBER: US 07/985,321  
 ; FILING DATE: 04-DEC-1992  
 ; APPLICATION NUMBER: US 07/429,791  
 ; FILING DATE: 31-OCT-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ingolia, Diane E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPND-01121  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2710 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 ; US-10-011-366-6

Query Match. 51.4%; Score 4458; DB 4; Length 2710;









QY 1437 I--ENSSDIQKIDHIGFNGEHRQYIFYSYIDNETKYNGFIDYSK-----KEGLFTAEPFN 1490  
 Db : : : : :  
 1803 TPEKINGSLEALEHNN-----IDAINSVIGLVQYARMKWNNDNISAIHDAG 1849  
 QY : : : : :  
 1491 E-SIIRNIYMPDSNNLFYSSKDLKDIRINKGVK-LLIGNYFDDMKVSLSFIEDTN 1548  
 Db : : : : :  
 1850 AVSDIKNIV--DKFLGGILTLNRRVYNGPGVSGASLEGFISGLEVCA-- 1897  
 QY : : : : :  
 1549 TYKLAGV---YLDENGVAQILKFMNAKSAKALNTSNNFLSINIKNIFYNLND-- 1600  
 Db : : : : :  
 1898 --RMGGTAGRYL--SNVAKVIL-----PLDII--GINISLYDSSLNHAHAKAT 1939  
 QY : : : : :  
 1601 PNIBFI---LDTNF-IIGSNSIGOFPE-----LICDKDKNIQPIYFNFKIKETSYLYV-- 1650  
 Db : : : : :  
 1940 TQIEVISTAIDVFSFSSINTALSIGAIAVPLAIA-----IVPITI-FSHEVKNVAVYVNO 1993  
 QY : : : : :  
 1651 -GNRQNLIVPEPSYHLD-----SGNISSTVINFSQKLYG---IDRYNVKUII 1694  
 Db : : : : :  
 1994 INERHKLMLAEKYLIDGSAKVLISINKATGIIDLSNNQVNLGNIYLDMPRENPPIL 2047  
 RESULT 12  
 US-09-134-001C-3159  
 ; Sequence 3159, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3159  
 ; LENGTH: 10182  
 ; TYPE: PRN  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3159

Query Match 4.0%; Score 348; DB 4; Length 10182;  
 Best Local Similarity 19.4%; Pred No. 1.9e-11;  
 Matches 354; Conservative 317; Mismatches 663; Indels 494; Gaps 97;  
 QY 20 QDEYVAILNALLEEYHNMSSESVKYLKLDINN-----LTDN-----YLN-- 61  
 Db : : : : :  
 7557 QKD---TILNHI--FSAPTRSQVGEKIASAKQLNMTKALRDSIADNNEILQSSKYFNE 7611  
 QY : : : : :  
 62 TYKSGRNKALKPKPEYLTMEVLELKNLSLTPV-----EKNLHFIIWIG 105  
 Db : : : : :  
 7612 SQQAYNQAVNKAHII-----NDQPTPMANDEIQSVLNEVVKQTKDLNHL---GD 7659  
 QY : : : : :  
 106 Q--IND-----TAINYNQKVDNSDYTVKFFYDSN-----AFLINTLKT 144  
 Db : : : : :  
 7660 QKLANDKTDQAOTIALNALYNLAQAQGNLETKVQ---NSNRFEQKVQLANQLNDAMKK 7716  
 QY : : : : :  
 145 IVESATNTLTSPRENLDNPEFYDKPKRMEIYDKOKHIDYKKSQIEE--NPPE-- 200  
 Db : : : : :  
 7717 LODALTGN--DAIKQTSYINEDTSQ-----QVNFDB---YTRGRKNIVAEQTPNMS 7765  
 QY : : : : :  
 201 -----IIDIINKYLSNEYSKDEALNKY---IEESLNKITANNGNDIRNLEKFADEDIV 252  
 Db : : : : :  
 7766 TNINTIADKI-----TEAKNDLHGQVKLQKQAQOQSINTI-----NQMTGLNQAQKEQL- 7813  
 QY : : : : :  
 253 RLYNSELVERWNLAASDLIRLSMLKEGQGVYLDVILPGIQDLFKSINKPDSITNTSW 312  
 Db : : : : :  
 7814 ---NOEIQOTRSEVHVQINKAQLNDS-----MNTLRQSTIDSEHYKQTSN 7858  
 QY : : : : :  
 313 EMIKL-----EAIMKYEYIPGVTSKNFMDLDEEVORSPESSALSSKSKSEIFLPLD 364  
 Db : : : : :

Db 7859 YINETVNGQTAYNNADVDRVKQIINQTSNPTWNPPL--EVERATSNVYKISK-DALHGERELN 7915  
 QY : : : : :  
 365 DIKUSPLEVKIAPANNV--INQALISUKDSVCSDL-----VINQIKNRVKIIN---- 411  
 Db : : : : :  
 7916 DNKSK-----TFVNVHLNLAQA---QKEALTHEIEQATVSVQNNIYNKAKALNNDMK 7967  
 QY : : : : :  
 412 -----DNLNPSINEGTDFNTTMKIFSDKLASISNEDNMFMFI-----KITNYL 454  
 Db : : : : :  
 7968 KLDIVAQODNVQRNNYINEDSTPQNMVND--TINHAQSIIDQVANTPMSHDEIENAI 8024  
 QY : : : : :  
 455 KVGFPADVRSTIN--LSGPGVTGAYQDILLMKDNSTNHLLEPELNRNEFP--KTKISQLT 512  
 Db : : : : :  
 8025 N-----NIKHAINALDGE-----HKLQQAQENA---NLLINSNDLNPQORDAINRLV 8069  
 QY : : : : :  
 513 E-----OBTISLWSFNQA-----RAKSOF---EYVKGVFECALED 546  
 Db : : : : :  
 8070 NEAQTREKVAQOLQAQALNDAMKHLNRSIQNSSVROESKYINASDADKKEQYNHAREV 8129  
 QY : : : : :  
 547 DNLDPAQNTVLDKQVSKILSSMKTRNKEYIHYIVQLQGDKISYEACN--LFSKDPYS 604  
 Db : : : : :  
 8130 ENIINEQHPTLDKEII-KQLTDGVNQANND-LNGVELLDADKQNAHQSIPTLMHLNQAQQ 8187  
 QY : : : : :  
 605 SILYQK-NIEGSETAYYYVADAEIKEDKYRIPYOISNKNRIKLTFIGHGKSEFTDTP 663  
 Db : : : : :  
 8188 NALNEKINNVAVTRTEVAAIIQQAQLLDHAMENLEESIKDEQVK-----QSSNYINEDSD 8242  
 QY : : : : :  
 664 ANLVDVSLSSSIETILN-----LAKADISPKYIEN-----LLGCNMFYSYIAEETVP 712  
 Db : : : : :  
 8243 VOETYDNADVHTTEILNQTVNPTLSIEDIEHAINEVNAQKQLRG-----KQKLYOTIDLA 8298  
 QY : : : : :  
 713 GKLLKIKDRVSELMPSISQDSITVSANQYEVRENEGR-----EILD 756  
 Db : : : : :  
 8259 DKELSKLDLTSQQSSSISNQIYTAKTTEVAQAIEKAKSLNHAMKALNKYKNADKVL 8358  
 QY : : : : :  
 757 HSGKWINKE-----ESIIKDISKEYISPNKENKIIIVSKYLHELSTLLQEIEN-- 807  
 Db : : : : :  
 8359 -SSRFINEQPEKKAYQQAINHVDISIHRQTNPEMDPTVINS-IHTELETAQNHLHGQK 8416  
 QY : : : : :  
 808 -----ANSSDIDLEKKVMLTECEIN-----VASNIDR-OIVEGRIBEAKNLT 848  
 Db : : : : :  
 8417 LAHAQQAANVINGLIHLNVAQREVMINTNATREKVAKNLDNAQALDKAMETLQQVV 8476  
 QY : : : : :  
 849 SDSINYIKNEFKLIESDSLYDLKHQGLDSDSHFISPEDISKTEGFRIPINKETGNS 908  
 Db : : : : :  
 8477 AHK--NNILNDSKYL--EDSKYQOQYDRVIADAGLLNQTTNPTLPKVDIVKQNV--- 8530  
 QY : : : : :  
 909 IFITEKEIFSEYATHISKEISNTKDTIFDNVNGKLVKKNLDAAEHVNTLNSAFFIOSL 968  
 Db : : : : :  
 8531 --LANEKILFG--AEKLSYDKSNANDEI-----KHNLYLNA----- 8563  
 QY : : : : :  
 969 LEYNTTKESLNSLVAMKVQ---VYAQLFSTGLNTITDASKVVELVSTAL----- 1015  
 Db : : : : :  
 8564 -OKOSIKDMI SHAALRTEVKQLLOQAQLDAMKSLEDKTQV--ITDITLPNYTBASDK 8621  
 QY : : : : :  
 1016 DETIDLLPTLSEGLPIIATII-DGVSLG--AAIKELSETNDPL---LRQIEAKI----- 1064  
 Db : : : : :  
 8622 KEKVD--QTVSHAQAIIDKINGSNVLDQVQALEQLTQASENLGDQDQVE-EAKVHAQ 8678  
 QY : : : : :  
 1065 ---GIMAVNLTAASTA--IVTSALGIASGFISILLVPLAGISAGISPLSVNNELILODKATK 1119  
 Db : : : : :  
 8679 TIDQLTHLNSLQQQTAKESVKNATKLEB-----IATVSNNAAQALNKVMGK 8723  
 QY : : : : :  
 1120 VIDFKHLSLAETEGAFLLDDKLIIMPQDDIVLSEIDFNNSNITLGKCEIWRAGSGSHT 1179  
 Db : : : : :  
 8724 LEQFINHADSVENSNDYRQADDDKIIAYDEALEHGQDIQKTNATQNETK----- 8772  
 QY : : : : :  
 1180 LTDDIDHFFSPSITIRKPMLSIYDVNLN- IKKEKIDPSKDLMLPNAPNRVFGYBGMWTP 1238  
 Db : : : : :  
 8773 -----QALQOLIYAETSINGFERLNHAPRALEYIKSLEKINNAAKSALEDKVTQSH 8824  
 QY : : : : :  
 1239 GFRSLD-----NDGTKLLDIRDHYEQFYWYFAFIADALITKLKPRYEDTNNRINLDGNT 1295  
 Db : : : : :  
 8825 DLLELEHIVNEGTNLNDIMGE-----LANAIVN-----NYAPTAKASIN----- 8862



[illegible]

```

RESULT 14
US-09-091-501B-10
; Sequence 10, Application US/09091501B
; Patent No. 6518413
; GENERAL INFORMATION:
; APPLICANT: Tinsley, Jonathon M
; APPLICANT: Davies, Kay E
; TITLE OF INVENTION: Utrophin gene expression
; FILE REFERENCE: 620-42
; CURRENT APPLICATION NUMBER: US/09/091,501B
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: PCT/GB96/03156
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: GB 9525962.8
; PRIOR FILING DATE: 1995-12-19
; PRIOR APPLICATION NUMBER: GB 9615797.9
; PRIOR FILING DATE: 1996-07-26
; PRIOR APPLICATION NUMBER: GB 9622174.2
; PRIOR FILING DATE: 1996-10-24
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 3433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (239) ... (250)
; OTHER INFORMATION: Description of Artificial Sequence: Full length
; OTHER INFORMATION: utrophin construct; Xaa = unknown
US-09-091-501B-10

```

Query Match 3.3%; Score 287; DB 4; Length 3433;  
 Best Local Similarity 18.6%; Pred. No. 2e-08;  
 Matches 318; Conservative 292; Mismatches 613; Indels 488; Gaps 85;

QY	27	ILNLEEEYHNMSSESVVEKYKLKIDNNLTNDVNTYTKSGRNKALKKPKKEYLTMEVL-E	85
DB	699	LLNWILKWTATOTTBIKYMKNQDTSSEMKKK-LKALEKEQERIRPRADELNQTGQILVE	757
QY	86	LKXNLSLTPVEKNLHFTWIGQINDTAINTYNOWKDVNS-----DYTVKFFYDVSNAF-	136
DB	758	QMKCEGLPTB-----BIKNVLEKVSSEKWNVSQHLDELERKIQLOEDINAYPKQL	807
QY	137	-----LINT-----LKKTIVESATNNLTSEFRE-----NL-----	161
DB	808	DELEKVIKTKEBWNVHTSISSESRQSLPSLKDSQRELTNLLGLHPKIEMARASCALMS	867
QY	162	--NDPEFDYNKF-----YKRMELIYDKQKHFIDYKQ-----TEENPEFIIDNI	205
DB	868	QPSAPDFVORGDFSLGRYQAOVQAEVDEHQHLENELKGPQHAYLETUKTLKDKLVNDSE	927
QY	206	IKTYLSNEYSKOLEALNKYIEE--SLNKITANN-----GNDIRMLEKFADEDLVRLYN	256
DB	928	NKAQVSLNVNLDAKVEKALQEKTYLDELEKQKPAHLKLAETKALEKNVHDPVEKLYK	987
QY	257	QELVE---RWN---LAAASDI-----LRISMLKEDGGV---YLDVDILPGIQDPLFKS	300
DB	988	QEFDDVOGKWNKLKVLVSXDLHLLEBIALTLRAFEADSTVIEKWD-----GVKDFLMQK	1042
QY	301	-----INKPSITYSWEMIKLEAIMK-----YKEYIP---GYT	331
DB	1043	QAAQGDAGLQRLQDCSAFVNEIETIESLKNMKIEITNLRSGPVAGIKTWQTRLGDY	1102
QY	332	SKNFOMLDEEV-----QRSFESALSSKDKSEIEFLPDDIKVSPLEVKIAPANNSVIN	384
DB	1103	QTQLEKLSKEIATQKRLSESQEKAAVLKDLAEQWMTQABEEYLERDFEYKSPEELE	1162
QY	385	QALISLKQSYCSDLVINQIKRYKILNDLNLPNSINEGTFDNTTMMKIFSDKLASINEDNM	444
DB	1163	SAVEEMKRAK-EDVLQKEV--RVKILKONI-----KLLAAKVPSSGGQELTS	1205
QY	445	MPWIKITNYL-----KVGFAPDVVRSTINLSGPGVYTGAYQDILLMFKD-NSTNIHLE	495
DB	1206	ELNVLENYQLLCNRIRGCHTLEEVWS-----CWTELHYLDLETTWLNTLE	1253
QY	496	PELRNPEFPKTKISQLTQOEITSLMSF--NQARAKSQFBEYKKGYPEGALGEDDNLDDFAQ	553
DB	1254	ERMSKEVLPKTDVANE-ALESLESLVRHPADNRTOIRELGOTLIDGG-----	1301
QY	554	NTVLDKDYVSKKILSSMKTRNEYIHYIVQLQGDKISYEASCNLFSDKDPYSSILYQKNIE	613
DB	1302	--ILD-DIISEK-LEAFNSRYEDLSHL---AESKQISLEKQLQVLRQETDQMLQVLESIG	1354
QY	614	GSETAYYYVADAIBKIDKIRIPYQIS-----NKNRIKUTFTGHGK	655
DB	1355	ELDKQLTTVLTD-----RIDAFQVPOBAQKIQAEISAHELTLEBLRNRMSQPLT---	1404
QY	656	SEFNTDTFANLDVDSLSSIEITILNKAADISPKYIEINLGCNMFSSYSIYAEETYPGKL	715
DB	1405	SPESTRAGSGQMDVLQRLKREVSTFKQFQKPAFQRMWLDCKRVLDGVKAE-----	1457
QY	716	LKIKDRVSELMPISQSDSITVSANQYEVIRINEEGKREILDHSGKWINKEESIHK---	772
DB	1458	-LHVLD-VKDVPDVLQTHLDKCMKLYTL--SEVKLEV-----ETVIKTRGHI	1502
QY	773	SKKEYISFNPKENKIIVKSYLHELSTLQEIIRNNANSDDIDLEKKVMLTECEINVASNI	832
DB	1503	VQKQOTD-NPKG-----MDBQLTSLKVLVNDLGAQVTEGQDLERASQLARKMKKEAASL	1556
QY	833	DRQIVGEBRTEAAKNLTSD-----SINYIKNEFKLIE-----SISDSLVDLKHONG	877
DB	1557	SEWLSATETELVQKSTSEGLGLDPTIEISWAKNVLDKLEKRAKADNTITISSAAL--QNL	1614

QY 878 LDDSHFISPEDISKTENGF-RIRFINKETGNSIF-IETEKEIFSYBATHISKEISNIKOT 935  
DB 1615 IEGSEPIERLVCNLNAGSRVTRWTDNCNLMHQNQLEIFDGNVAHISTWLQ-AEA 1673  
QY 936 IFDNVNGK-----LVKKV--NLDAH--EVTNLSAFFIOS-----LIEY 971  
DB 1674 LLDEIEKPTSKQEEIVKRLVSELDQANLQVENVRDQALILMARGSSRELVEPKLAE 1733  
QY 972 NTKKESLS-----NLSVAM-----KVQVYAOFLPSTCL-----NTTIDKAVVE--L 1010  
DB 1734 NNFEEKVSHIKSAKLLIAQEFYQCLVTFETFGVPFSDLEKLENDIENMLKPEVEXHL 1793  
QY 1011 VSTALDETDLPTLSEGLPIIATIIDVSLGAALKELSETNDPLLRQIE-----AKIGI 1066  
DB 1794 ESDEDEKVD-----EESAQIEEVLORGEMLHQPMENKKEKIRL 1834  
QY 1067 MAVNLTAASTAIVTSAL-----GIASGPSILLVPL-----AGISAGIPS 1105  
DB 1835 QLLLLHTRYNKIKAIPIQKQKMGQASGIRSSLLPTDYLVEINKILLCMDDVELSLNVPE 1894  
QY 1106 LVNNELILOKATKVIDYFKHISLAETEGAFLLDDKIIM---PQDDLVLSEIDFNNSI 1162  
DB 1895 L--NTAIYEDFS-----FQEDSLKNIKDQDLQGEQIAVIEHKQPDVIL----- 1936  
QY 1163 TLGKCEIWAEGGSG---HTLT-----DDIDHPFSSPSITY---RKPWLSIYDVLIK 1209  
DB 1937 -----EASGEAIQIBDTLTQLNKWDNRINMYSDRKGCFDRAMEWRQFHCOLDNL 1988  
QY 1210 KKKIDPSKOLMVLNPNP-----RVFGYEM-----CWTGFRSLDNDGKLLDRIRD 1256  
DB 1989 TQWITEABELLVDTCPAGGSLDLKARIHQOELEVGISSHQPSFAALNRTGDIQVKL-S 2047  
QY 1257 HVEGQPYWYFAFIA---DALITKLPRYEDTVNRLNDCNTRSPFIVPITTEQIRKNLS 1313  
DB 2048 QADGSPKLKLAGLQNRWDIAVAEVKDR-----QPLRGESK-----QVMK--- 2088  
QY 1314 YSFYSGSGSYSLSPYNNIDNLNVENDTWIDVNVV-KNITIE-SDEIQKGLIENI 1371  
DB 2089 -----YRHLQD-----EICWLTKAHAHMKRSTTELGENLOE---LRDL 2125  
QY 1372 LSKLNIEDNKI-ILNHNHTNFYGD-----INESNRFISLFTS-ILEDI-----NI 1414  
DB 2126 TQEMEVAEKLWLNRTELEMLSDKLSLPERDKISELSRTVMNTWNIKREBVPITLKEC 2185  
QY 1415 IIEIDLVSYSKILLSGCNWK--LIENSSDI 1443  
DB 2186 IQEPSSVSQT-RIAAHPNVQKVLVSSASDI 2215

## RESULT 15

US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 5080

LENGTH: 3696

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5080

Query Match 3.2%; Score 281; DB 4; Length 3696;  
Best Local Similarity 18.0%; Pred. No. 5.1e-08;

Matches 335; Conservative 284; Mismatches 693; Indels 546; Gaps 79;  
QY 48 KKDINNLNTNLYATYKKSGRKAL--KKPKXYLTMEVLELKNLSLTPVKNLHFIWIGG 105  
DB 513 KLRPN-----ISTPREVTNEAIKYTFSEY-----YINTNDMTVT-----G 550  
QY 106 QINDTAIYNQWQDVNSDYTVKVFYDSNAFL-----INTLK-----KTIVESATNNLTBSF 157  
DB 551 QOTFPSINVNN-KDDLSEQVNDKIIIPSNYTLASVKNYKLERAQTVLDBETNNT--PF 607  
QY 158 RENLNDPFD-----YKPYRKRMEIYYQKHFIIDYKYSQIE--ENPEFIIDN 204  
DB 608 NORYSQTOIDLLHQLTTLINRVASRE--INDKAQEMTDVAVDSTELTTEKOTLVDQ 665  
QY 205 I--IKTYSLSYKDL-----EALNKYIEESLNKITANNNDIRNLEKFADEDLV 252  
DB 666 IENHKEISNNIDDELTDGGERVKEAGLTLESDTPHPVTKPNARQVNNRADQOKTLI 725  
QY 253 RLYNOELVERWNLAA-----ASDIL-----RISMLKEDGGVYLDVILPGIOPD 296  
DB 726 RNNHEATTEEQNEAIRQVEAHSSDAIAKIGAEATDVTTVNEARDNGTKLIATDV-----PN 780  
QY 297 LFKSINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFSAKSKSDK 356  
DB 781 PTKAARAAAVTNGANSKIK-----LSGPGVYTGAYQDLLMFKD-----NSTN 490  
QY 357 SEIFLPDDIKVSPLEVKIAFNNSVINQALISLKSYSVDLVNQIKRKYKILNDLNP 416  
DB 816 -----AIALVNRS-----KDEAIQN--INTAGQ-----NDDVTE 842  
QY 417 SINEGTDFTNTWKI-----FSDKLASI--SNEDNMFMKIKITNYLKVGFADPVRSTIN-- 467  
DB 843 AQNNGT--NTIQVPLTPVKQNAIATINAKADEQKRLIQANNNATTEKADAERKVEA 900  
QY 468 -----LSGPGVYTGAYQDLLMFKD-----NSTN 490  
DB 901 VITANQINATTNRDQVDAQTTGSGIISAIISPAATKIKEDARAAVEAKAIQNOQINSNN 960  
QY 491 IHLLEPELRNFPFKTKISQLTQEITSLWSFNQARAKSQFEYKKGFEALGEDDNL 550  
DB 961 MATTEEK-----EDALNQVEAHKQAAIATINQAOQSTQVSEAKNGNINTINQOPNAV 1013  
QY 551 FAQN--TVLQDYVSKIL-----SSMKTRNKYIHYIVQLQGDKISYEASCNLFSDOPY 603  
DB 1014 KNWTKITILEQKNEKSAIAQTPDATTEEQEAUSAVSQAVTNGITHINQAN--SNDV 1071  
QY 604 SSILYQK-----NIEGSETAYYYVADAIEKEI-----DKYRIPYQISNK 643  
DB 1072 DQELSNAEQIITHNVNVQKQPQARQALIAKTNEKQSAINSNDEGTIEBEKQAIQSLNDA 1131  
QY 644 RNKLTPIFGHKSSEPTDTPANLDVDSLSSSEIETILNLAKAD-----ISPKYIEIN 694  
DB 1132 KNLADEQITQASNQNVNDAINIGISNT-SKIQT--NFTKKQOARDQVQKQFEKAEALN 1188  
QY 695 LIGCMPSYSIVAEBETYPGKLLIKDRVSELMPISQ-----DSITVSA-----NQ 741  
DB 1189 -----STPATQDEKQDALTRLTOAKETALNDINQATNQNVDTALTSGIQNIQTQ 1240  
QY 742 YEVRINEEGKEILD-----HSGKWINKEESIIKD-----ISKEYISFPN 782  
DB 1241 VNVKQKEAKTTINDIVQOHKQSIQNDDDATTEKEVANLNVASQQNVISKIDNATTNN 1300  
QY 783 KENKIIVKSKYLHELSTLLOEIRNANNSSDIDL---EKKVMLTECEINVASNIDQIVEG 839  
DB 1301 QIDGIVSGROSINAITPDTSIKRNA-KNDIDIKAAKKIKIQ--RINDATDEETIQEANR 1357  
QY 840 RIEEAK-----NLTSDSINYIK-NEFKLIISI-----SDSLYDLKHQGLDSDH 882  
DB 1358 KIEEAKIEAKONIQRNSTFDQVNEAKTNGINKIENITPATTVKSEARQAVQNKANEQINH 1417  
QY 883 FISFEDISKTENGFRIEINKETGNSIFIEETEKEIFSYBATHISKEISNIKDTIFDNVNG 942  
DB 1418 IQNTPDATNEEKQEAIRNVSAE-----LARVQAQINAEHTTQGVKTIKDDAITSLSRINA 1472

Search completed: November 5, 2003, 19:33:30  
Job time : 33.75 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: November 5, 2003, 19:27:13 ; Search time 23.125 Seconds  
(without alignments)  
7069.698 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1700  
Perfect score: 8677  
Sequence: 1 MNLVNAQLQKVVYKFRIQ.....YLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8646	99.6	2364	2 I40884	Cytotoxin L - Clos
2	6733.5	77.6	2366	2 S10317	toxin B - Clostrid
3	6711	77.3	2367	2 S70172	toxin A - Clostrid
4	4458	51.4	2710	2 A37052	alpha-toxin - Clos
5	2108.5	24.3	2178	2 S55805	toxin B - Escherich
6	496.5	5.7	3169	2 T00296	rhoctry protein -
7	470.5	5.4	2401	2 T28676	rhoctry protein -
8	437	5.0	2269	2 T28677	hypothetical prote
9	424	4.9	2166	2 G70163	hypothetical prote
10	423.5	4.9	4688	2 F82885	adherence factor T
11	403	4.6	3225	2 D81702	hypothetical prote
12	398	4.6	1979	2 C71622	hypothetical prote
13	390	4.5	3724	2 T18427	hypothetical prote
14	386.5	4.5	3394	2 T18501	hypothetical prote
15	370	4.3	1711	2 T18429	hypothetical prote
16	359	4.1	1939	2 T18372	repeat organellar
17	356.5	4.1	4981	2 T18489	reticulocyte-bind
18	351	4.0	2829	2 A42771	adherence factor T
19	348.5	4.0	3335	2 H81702	ORF MSV156 hypothe
20	345.5	4.0	1127	2 T28317	conserved hypothet
21	342	3.9	2819	2 A90551	hypothetical prote
22	340	3.9	5005	2 F82884	serine/threonine-s
23	337	3.9	2485	1 H71621	asparagine/asparta
24	330.5	3.8	3844	2 T18402	adherence factor T
25	330	3.8	3255	2 G81702	hypothetical prote
26	327.5	3.8	4550	2 T18440	DNA-directed RNA p
27	327	3.8	2910	2 T28156	major merozoite su
28	318.5	3.7	1701	2 A26868	hypothetical prote
29	318	3.7	3973	2 B71612	

30	317.5	3.7	1726	1 SAZQGM	major merozoite su
31	317.5	3.7	1726	2 A45948	major merozoite su
32	317.5	3.7	2599	2 F90608	ABC transporter pe
33	316.5	3.6	1701	2 A54498	major merozoite su
34	316	3.6	1447	2 F82909	hypothetical prote
35	314.5	3.6	2496	2 A71616	secreted protein P
36	314	3.6	1819	2 D97033	uncharacterized pr
37	313	3.6	1306	2 T28313	ORF MSV152 probabl
38	311	3.6	1302	1 JC6009	surface-located me
39	310.5	3.6	1830	2 E82909	conserved hypothet
40	309.5	3.6	2244	2 F90563	hypothetical prote
41	309	3.6	1639	2 S05603	major merozoite su
42	306.5	3.5	1516	2 E71619	RAD2 endonuclease
43	306.5	3.5	2136	2 A05037	hypothetical prote
44	305.5	3.5	1957	2 T38077	hypothetical coile
45	303.5	3.5	1802	2 G71616	hypothetical prote

ALIGNMENTS

RESULT 1

I40884  
cytotoxin L - Clostridium sordellii  
C:Species: Clostridium sordellii  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40884  
R:Green, G.A.; Schue, V.; Montell, H.  
Gene 161, 57-61, 1995  
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii  
A:Reference number: I40884; MUID:95369733; PMID:7642137  
A:Accession: I40884  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2364 <RES>  
A:Cross-references: EMBL:X82638; NID:g1000694; PIDN:CAA57959.1; PID:g1000695  
C:Superfamily: cpl repeat homology  
C:Keywords: cytotoxin

Query Match 99.6%; Score 8646; DB 2; Length 2364;  
Best Local Similarity 99.8%; Pred. No. 1.2e-303;  
Matches 1696; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	MNLVNAQLQKVVYKFRIQDEYVAI	NALALEEYHNMSSESVVEKYLKLDI	NNLT	DNYL	60
Db	1	MNLVNAQLQKVVYKFRIQDEYVAI	NALALEEYHNMSSESVVEKYLKLDI	NNLT	DNYL	60
Qy	61	NTYKSGRNKALKKPKYLTWEVLE	LNKNSLTPVEKNLHFIWIGQINDT	AINYINQ	WKD	120
Db	61	NTYKSGRNKALKKPKYLTWEVLE	LNKNSLTPVEKNLHFIWIGQINDT	AINYINQ	WKD	120
Qy	121	VNSDYTVKVVDSNAFLNTLTKT	IVESATNTLTSPFENLNDPFDY	NKFKRMEI	IY	180
Db	121	VNSDYTVKVVDSNAFLNTLTKT	IVESATNTLTSPFENLNDPFDY	NKFKRMEI	IY	180
Qy	181	DKQKHFDIYKSOIENPEFI	DNIIKYLSNEYSKOLEALNKY	IEESLNK	TANNNDI	240
Db	181	DKQKHFDIYKSOIENPEFI	DNIIKYLSNEYSKOLEALNKY	IEESLNK	TANNNDI	240
Qy	241	RNLEKFADEDLVRLYNQEL	VERWNLAAASDILRISMLKED	GGVYLDV	ILPGIQDLP	FKS 300
Db	241	RNLEKFADEDLVRLYNQEL	VERWNLAAASDILRISMLKED	GGVYLDV	ILPGIQDLP	FKS 300
Qy	301	INKPDSINTSWEMIKLEAI	MKYKEYIPGYTSKNPMDLDEE	VQSFESAL	SKSKSEIF	360
Db	301	INKPDSINTSWEMIKLEAI	MKYKEYIPGYTSKNPMDLDEE	VQSFESAL	SKSKSEIF	360
Qy	361	LPDDDIKVSPLVKIAPANN	SVINQALISLKDSYCDLVINO	IKNRYKIL	NLNLP	SINE 420
Db	361	LPDDDIKVSPLVKIAPANN	SVINQALISLKDSYCDLVINO	IKNRYKIL	NLNLP	SINE 420
Qy	421	GTDFNTTKIFSDKLASIS	SNEDNMFMIKITNYLKV	GAPDVR	STINISG	PGVYTGAYD 480
Db	421	GTDFNTTKIFSDKLASIS	SNEDNMFMIKITNYLKV	GAPDVR	STINISG	PGVYTGAYD 480

Db 421 GTDFNTMTKIFSDKCLASISNEDNMFMKIITWLVKGFAPDVRSTINLSGPGVYTCAYQD 480  
Qy 481 LLMFKDNSTNIHLPELRFNFPKTKISQLEQETISLWSFNQARAKSQFEYKKGVE 540  
Db 481 LLMFKDNSTNIHLPELRFNFPKTKISQLEQETISLWSFNQARAKSQFEYKKGVE 540  
Qy 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIYVQLQGDKISYEASCNLFPSK 600  
Db 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIYVQLQGDKISYEASCNLFPSK 600  
Qy 601 DYPSSILYQKNEGSETAYVYVADAIEIKEDIKYRIPYOISNKRNIKLTFFIGHGSEFNT 660  
Db 601 DYPSSILYQKNEGSETAYVYVADAIEIKEDIKYRIPYOISNKRNIKLTFFIGHGSEFNT 660  
Qy 661 DTFANLDVDSLSEETILNLAKADISPKYIEINLLGCNMFYSIYASETYPGKLLKIK 720  
Db 661 DTFANLDVDSLSEETILNLAKADISPKYIEINLLGCNMFYSIYASETYPGKLLKIK 720  
Qy 721 DRVSELMPSISQDSITVSANQYEVRIENERGKEILDHSGKWNKESIIKDISKEYISF 780  
Db 721 DRVSELMPSISQDSITVSANQYEVRIENERGKEILDHSGKWNKESIIKDISKEYISF 780  
Qy 781 NPENKIIYKSKYLHELSTLLQEIERNANSSDIDLEKKVMTFECINVASNIDRQIVEGR 840  
Db 781 NPENKIIYKSKYLHELSTLLQEIERNANSSDIDLEKKVMTFECINVASNIDRQIVEGR 840  
Qy 841 IEBAKNLTSDSINYIKNEPKLESISDSLYDLKHQGLDDSHFISFEDISKTEGFRIRF 900  
Db 841 IEBAKNLTSDSINYIKNEPKLESISDSLYDLKHQGLDDSHFISFEDISKTEGFRIRF 900  
Qy 901 INKETGNSIFITEKEIFSEYATHISKEISNIDKTIFFONVNGKLVKVNLDAAHEVNTLN 960  
Db 901 INKETGNSIFITEKEIFSEYATHISKEISNIDKTIFFONVNGKLVKVNLDAAHEVNTLN 960  
Qy 961 SAFFIQSLLEYNTKESLNSLVAMKVQVYQALFSTGLNTITDASKVELVSTALDETID 1020  
Db 961 SAFFIQSLLEYNTKESLNSLVAMKVQVYQALFSTGLNTITDASKVELVSTALDETID 1020  
Qy 1021 LAPTISEGPIIATIDGVSIGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAVT 1080  
Db 1021 LAPTISEGPIIATIDGVSIGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAATAVT 1080  
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELLQKATKVIDYFKHISLAETEGAFITLD 1140  
Db 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELLQKATKVIDYFKHISLAETEGAFITLD 1140  
Qy 1141 DKIIIMPQDLVLSIEDFNNSITLGKCEIWRAGSGGHTLTDDIDHFFSSPSITVRKPKWL 1200  
Db 1141 DKIIIMPQDLVLSIEDFNNSITLGKCEIWRAGSGGHTLTDDIDHFFSSPSITVRKPKWL 1200  
Qy 1201 SIYDVNLIIKKEKIDFSKOLMVLNAPNRVFGYEMGWTGFRSLDNDGTKLLDIRDHVEG 1260  
Db 1201 SIYDVNLIIKKEKIDFSKOLMVLNAPNRVFGYEMGWTGFRSLDNDGTKLLDIRDHVEG 1260  
Qy 1261 QYWRVYFAIADALITKLPRVEDTNRINLGNTRSFTVPVITTEQIRKNLSYFYSGS 1320  
Db 1261 QYWRVYFAIADALITKLPRVEDTNRINLGNTRSFTVPVITTEQIRKNLSYFYSGS 1320  
Qy 1321 GSYSLSPYNNMIDLNVENDTWIDVDNVVKNITTESDEIQKGELENIILSKLNIEDN 1380  
Db 1321 GSYSLSPYNNMIDLNVENDTWIDVDNVVKNITTESDEIQKGELENIILSKLNIEDN 1380  
Qy 1381 KIILNNHTNFGDINESNRFSLTFSILEDNIIEIDLVSYSKILLSGNCKLIENS 1440  
Db 1381 KIILNNHTNFGDINESNRFSLTFSILEDNIIEIDLVSYSKILLSGNCKLIENS 1440  
Qy 1441 SDIQKIDHIGFNGEHOKYIFYSIDNETKYNGFDYSKKEGLFAEFSNESIIRNIYMP 1500  
Db 1441 SDIQKIDHIGFNGEHOKYIFYSIDNETKYNGFDYSKKEGLFAEFSNESIIRNIYMP 1500  
Qy 1501 DSNLPIYSSKOLKDIRIINKGDKVLLIGNYFKDDMKVSLSTPTIEDTNTIKLVGYLDEN 1560  
Db 1501 DSNLPIYSSKOLKDIRIINKGDKVLLIGNYFKDDMKVSLSTPTIEDTNTIKLVGYLDEN 1560

Qy 1561 GVAQILKPMNAKALNTSNLMNFLESINIKNIFYNNLOPNIBFILDNTFIISGNSIG 1620  
Db 1561 GVAQILKPMNAKALNTSNLMNFLESINIKNIFYNNLOPNIBFILDNTFIISGNSIG 1620  
Qy 1621 QFELICDKDKNIQYFIFNFKIKETSYTLVGNRQNLIVPEPSYHLDDSGNISSTVINFSQK 1680  
Db 1621 QFELICDKDKNIQYFIFNFKIKETSYTLVGNRQNLIVPEPSYHLDDSGNISSTVINFSQK 1680  
Qy 1681 YLYGIDRYVNVKVIITAPNLYT 1700  
Db 1681 YLYGIDRYVNVKVIITAPNLYT 1700

## RESULT 2

SI0317

toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_change 15-Oct-1999

C:Accession: SI0317; S21894; S22434

R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.

Nucleic Acids Res. 18, 4004, 1990

A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.

A:Reference number: SI0317; MUID:90326540; PMID:2374729

A:Accession: SI0317

A&gt;Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-2366 &lt;BAR&gt;

A:Cross-references: EMBL:X5138; NID:g40442; PIDN:CAA37298.1; PID:g40443

R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, N.

submitted to the EMBL Data Library, July 1991

A:Description: Comparative analysis of Clostridium difficile toxins A and B.

A:Reference number: S21894

A:Accession: S21894

A:Molecule type: DNA

A:Residues: 1271-2366 &lt;EIC&gt;

A:Cross-references: EMBL:X60984; NID:g40445; PIDN:CAA43299.1; PID:g40446

R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, N.

Mol. Gen. Genet. 233, 260-268, 1992

A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.

A:Reference number: S22434; MUID:92293124; PMID:1603068

A:Accession: S22434

A:Molecule type: DNA

A:Residues: 1791-2366 &lt;VON&gt;

A:Cross-references: EMBL:X60984

C:Genetics:

A:Gene: toxB

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 77.6%; Score 6733.5; DB 2; Length 2366;  
Best Local Similarity 76.4%; Pred. No. 8.4e-235;  
Matches 1299; Conservative 190; Mismatches 210; Indels 1; Gaps 1;

Qy 1 MNLVNAQLOKQVVKPRIQDEYVAILNALEEYHNNESSVWEKYLKLDINLTDNVL 60  
Db 1 MSLVNRKQLEKMANVRFTQDEYVAILDALEEVHNSENTVVEKYLKLDINSLTDIYI 60  
Qy 61 NTYKSGRNKALKKPKLEYLTMEVLELKNNSLTPVEKNLHFIIWGGQINDTAIYNQWKD 120  
Db 61 DTYKSGRNKALKKPKLEYLTMEVLELKNNSLTPVEKNLHFIIWGGQINDTAIYNQWKD 120  
Qy 121 VNSDYTVKVVDSNAFLINTLKTYIVESATNTLTLESFRENLDNDFDYNKPKRKMELIY 180  
Db 121 VNSDYNNVNFVDSNAFLINTLKTYIVESAINDTLTLESFRENLDNDFDYNKPKRKMELIY 180  
Qy 181 DKQHFIDYKKSQLEENPEFIIDNIITKYLISNEYSKOLEALNKYIEESINKITANGNDI 240  
Db 181 DKQKNFINYKQAREENPELLIDDIIVKYLISNEYSKEIDELNTYIEESINKITQNSGNDV 240  
Qy 241 RNLKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPKS 300  
Db 241 RNFBEFKNGESFNLYEQSLVERWNLAASDILRISALKEIGWYLDVDMPLPGIQDLPFS 300







A;Molecule type: DNA  
A;Residues: 1-55,'V',57-2079,'L',2081-2549,'S',2551-2710 <SAU>  
A;Cross-references: EMBL:X51797

Db 1599 NINFDVKYFTLVGKTNLGVFEICDNNKNDIYFGWKTSSTSSKSTIPSGNGRNVVVERP 1658

Qy 1662 YHLDSCNLSSTVINFSQKLYGIDRYVKNVLIAPNLYT 1700

Db 1659 YN-PDTGEDISTSLDFSYPYLYGIDRYVKNVLIAPDLYT 1696

RESULT 5

S55805

alpha-toxin - Clostridium novyi (ATCC 19402)

C:Species: Clostridium novyi

A:Variety: ATCC 19402

C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999

C:Accession: S55805; S71294; S71158; S44273; I40834; S44272

R:Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.

Mol. Gen. Genet. 247, 670-679, 1995

A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium novyi

A:Reference number: I40834; MUID:95342160; PMID:7616958

A:Accession: S55805

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-2178 <HOS>

A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724

A:Accession: S71294

A:Molecule type: protein

A:Residues: 1-15 <HOS>

R:Hofmann, F.

submitted to the EMBL Data Library, March 1995

A:Reference number: S71158

A:Accession: S71158

A:Molecule type: DNA

A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>

A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724

R:Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.

submitted to the EMBL Data Library, July 1993

A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the family Clostridiaceae

A:Reference number: S44272

A:Accession: S44273

A:Molecule type: DNA

A:Residues: 1-243; I204-2178 <HOA>

A:Cross-references: EMBL:Z23281

C:Genetics:

A:Gene: tcn-alpha

C:Superfamily: cpl repeat homology

C:Keywords: virulence factor

F:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 24.3%; Score 2108.5; DB 2; Length 2178;

Best Local Similarity 31.2%; Pred. No. 2 5e-68;

Matches 551; Conservative 358; Mismatches 708; Indels 147; Gaps 46;

Qy 3 LVNKAQLOKVVYKFRIOEDYVAILNALAEYH-NMSESSVVEKYKLDKINNLDNYLN 61

Db 2 LITREQLMKIASIPKRPPEYNILDALENFRDIEGTSVKEIYSKLSKLNELVDNYQT 61

Qy 62 TYKSGRNLKALKFKKYLWELVLEKNSLTP--VEKNLHFWIGGQINDTAINYNQWK 119

Db 62 KYPSSGRNLALFNDRSLYSSELRELKNGRTSTIASKLSFWIGGPISDQSLSEYNNMK 121

Qy 120 DVNSDYTVKPVVDNSAFLNTLKTIVESATNNTLESFRENLDNPFYDYNKFRKRETI 179

Db 122 MFNKNIRLFPYDKNSLLVNTLKTALIQESSKVIIEQNSNILDGTGYHNKYSYDRMKUI 181

Qy 180 YDKQHFIDYKSIQIEENPEFIIDNIKTLYSNEYSKOLEALNKYIEESLNKITAANNND 239

Db 182 YRKRELKMLYENMKQNS--VDDIIINFLSNYFYKIDGLKNQKNNNNKMIAGATD 238

Qy 240 IRLNLEFAEDLVRLYNQELVERWNLAASDILIRSMLEKGGVYLDVILPGIQDLPK 299

Db 239 I-NTENILNTKLSYQELIQTNNLAASDILIRAILKKGVCYCDLDFPGVNLNLFN 297

Qy 300 SINKPDSITNTSMWIKLEAIMKYEYIPGYTSKNFMDLDEEVQSFESALSSKSKSEI 359

Db 298 DISKPNMGDSNYWEAAIFEAJANEKLMNNPYKYTMEQVPSSEIKERILSVRN-HDINDL 356

Qy 360 FLPLDDIKVSPLEV-----KJAFANNSVINQALSLKSDSYCSDLVINOIKRYKILNDNL 414

Db 357 ILPLGDIKISQLEILLRLKAATGKTKFSNFIISNNDSLTLNLLISQLENYEILNSII 416

Qy 415 NP--SINEGTD--FNTTMKIFSDKLASISNDNMFMKITNYLKVGFAPDVRSTINLSG 470

Db 417 QEKPKICETYDSYNSVSELVLETPKNLSWDGSSFYQOIIGYSSGPKPEVNSIVFSG 476

Qy 471 PGVYTGAQDILLMFKDN-----STNIHLLEPELRNFFPKTKISQLTEQBITSLWSFNQ 524

Db 477 PNIYSATCDTVHFIKNTFDMSSNQEQIFEAS--NNLYFSK-----THDEFKSSWLLRS 529

Qy 525 ARAKSQPEYKGYFEGALGDNDLDPQNTVLDKDYVSKKI--LSSMKTRNKEIYHIV 582

Db 530 NIAKEFOKLAKTYIGRTLANTYEDGLNFNKRVRTTSSELLKVIIEVNSTKIYENYDLNML 589

Qy 583 QLOQDKISYEASCNLFSDPYSSILYQKNIEG-SETAYVYVADAEIKEDKRYPIYOIS 641

Db 590 QIQGDDISYSAVNVFGKNPKSIL----IQGVDDFANVFYFENGIVOSDNNINLSRPN 645

Qy 642 NKRNILKTLFIHGKSEFNTDTFANLDVDSLSEI-----ETILNLAKADISPKYIEINLL 696

Db 646 DIKKIKLTLIGHGENVFNPKLFGGKTVNDLYNIIKPKQLHLLREGVILKNKYLKINIL 705

Qy 697 GGNMFSYIABETYPGKLLKIKDRVSELMP-SISQDSITVSANQYEVNRNEEGREIL 755

Db 706 GCMFTPKVDINSTPVGKLFNKIS---RDLPQKGFKNQLEISANKYAIRINREGREVL 762

Qy 756 DHSGKWINKERSIIKDISKEYISPNKPKENKIIVKSKYLHELSTLQIRNNANSSDIDL 815

Db 763 DYFGKWNSTDLIAEQISNKVTVWNEVENTLSARVEQLNKVAEPKDI-----NSIIQT 817

Qy 816 EKKVMTLCEINVASNIDRQVEGRIBEAKNLTSDSI----NYIKNEFKLIESISDSLYD 871

Db 818 TNNQELKQSLVNTVADLTLYSELKEDIPELDNIQIKERILLNEISRLHDFSNIL 877

Qy 872 LKHQGLDDSHFISFEDISKTEGFRIRFINKETGNSIFETEKEIFSEYATHISKELSN 931

Db 878 FYQKNISNNMIILFDSIIKEKDYNNVKNANKITGETSVIKTYSDSLWNFTNKYKIVDD 937

Qy 932 IKDTIFDNVNGKLVKKNVLDAAHEVNTLNSAFFIQSLJEYNTTKESNLVSAMKVQVYA 991

Db 938 IKGIIVKDIINGEFIKKADFIEQNPSPNSLNSAMLMQLLDYKPYTEILTNTMTSLKQVYA 997

Qy 992 QLFSTGLNTITDASKVVELSTALDETIDLLPTLSEGLPIIATIIDGVSGLGAIKELSET 1051

Db 998 QIFOLSIGAIQEAITEVITIIISDALNANFNILSKLVGSSVASVIIIDGINLIAALTELKNV 1057

Qy 1052 NDPLLRQIEIAKIGIMAVNLTAATAIVTSALG---IASGFSILLVPLAGISAGIPLSVN 1108

Db 1058 KTNFERKLEIAKGVMSYIGFLESLLSGLLGATAVSEILGVIISVPVAGLVGLPSLVN 1117

Qy 1109 NELLIQDKATVIDYFKHISLAETEGATLDDKTIIMPODDLVLEISIDFNNSNITLGKE 1168

Db 1118 NILVGEKYNQILDYFSKPYIVGKNPFS--IQDNIIIPYDDIAITELNFKNYKFKYGA 1176

Qy 1169 IWRABGGSGHTLTDIDHFFSSPSITYRKPMLSIYDVLNIIKKEKIDFSDKMLVLPNAPNR 1228

Db 1177 ISGFEKSGHTVGNIDHVSAPSILDHIE--LSIYPALKLNDTNLP-KGNVLLPSGLNK 1234

Qy 1229 VFGYEMGWTPGFRSLDNDGTLLDRIDHY-----EGOFYWRYPAFIADALITKUKPRYE 1283

Db 1235 VYKPEISAAGANSQEGNGVEVLNIRNYVDNSNGNTKFPWKYEAPP--EYSPSYMRVEYF 1293

Qy 1284 DTNVRINDGNTRSPIVPVITTEQIRKNLSYFSGSGSYLSLSPYNNNIDNLVENDT 1343

Db 1294 DTKNVILDNENKTLIIPVLTIDENRNKISVEILGGQYVILPWPQNTNINIVSKNDI 1353

Qy 1344 WVIDVNVVQKITTESDEIQKGELIENLSKLNIEDNKIILNNHTINFPYGDINESNRFTS 1403

Db 1354 MNFVSVYIVKESKIEDNKVFLDGFINNIFSTLKVSDGFKIGKQFIS-----IKNTPRAIN 1409

Qy 1404 LTFSEILEDINIIIEIDL-VSKSVKILLSCNCKMLIENSSD-IQOKIDHIGFNGEHQKYPF 1461

Db 1410 LSKFINNI-VIVSVYLNHEKSNSTIISDLDNDKNFNDLNDINYLGLSGISDNTIN 1468

Qy 1462 YSIDNETKNGFIDYSKEGLFTAFPSNESIRNIYMPDSSNNLFIYSGSKOLKDIRIINK 1521

Db 1469 CIVRNDVYMECKI-----FLNEK-KLVFIQNELELHLYDS-----VNK 1506

Qy 1522 GVDKLLIGN-----YFKDDMKVSLSTFTIEDT-----NTIKLNGVYLDENGVAQI 1565

Db 1507 -DSQYLNNPINNVKYGKGYVEGTFLINSTENKYSLYIENKMKLGLYL-ESSVFKT 1564

Qy 1566 LKFMNAKSAIANT-SNSLMNPLESINIK--NIFYNNLDPN--IEFILDYN--FIISGS 1616

Db 1565 IQDKYSEKENVYILSLIKKFTVNIQCPMWISGVGDENRYLEYMLSTNKKIING- 1623

Qy 1617 NSIQFELICDKDKNIQPYFINFKISETSYTLVGNRQNLIVEPSYHLDDSGNISSTVIN 1676

Db 1624 ---GYWE---ND-----FNNYKIVDF-----EKCNNVIVSGSKNLNSEGDLADT-ID 1662

Qy 1677 FSQKLYGLDRVNVKVIAPNLYT 1700

Db 1663 VLDKOLENL--YIDSVIIIPKPYT 1684

RESULT 6

T00296

toxin B - Escherichia coli plasmid p0157

C:Species: Escherichia coli

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 17-Nov-2000

A:Accession: T00296; T42195

R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yateudo, H.C.; Kubota, S.; Shinagawa, H.

DNA Res. 5, 1-9, 1998

A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli serotype O157:H7.

A:Reference number: Z14127; MUID:98290540; PMID:9628576

A:Accession: T00296

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3169 <MA>

A:Cross-references: EMBL:AB011549; NID:G4589740; PIDN:BAA31815.1; PID:G3337056

A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952

R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.

Nucleic Acids Res. 26, 4196-4204, 1998

A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli serotype O157:H7.

A:Reference number: Z22068; MUID:98391744; PMID:9722640

A:Accession: T42195

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-236, D', 238-1887, 'I', 1889-3169 <BUR>

A:Cross-references: EMBL:AF074613; PIDN:AAAC70163.1

A:Experimental source: strain EDL933; serotype O157:H7

C:Genetics:

A:Gene: toxB

A:Genome: plasmid p0157

A:Note: L7095

C:Superfamily: Escherichia coli probable cytotoxin

C:Keywords: cytotoxin

Query Match 5.7%; Score 496.5; DB 2; Length 3169;

Best Local Similarity 20.0%; Pred. No. 4.4e-10;

Matches 419; Conservative 336; Mismatches 719; Indels 621; Gaps 105;

Qy 14 YVKFRI---QDEYVAINAL--EYHNHSSSVSEVKEVYKLDKIDNNLTNYNTYKSGR 68

Db 160 YIKIRTRGAEDQTITITQSLINELLNGVDRNTI-PFQKISELNDIHSYENMQIKSR 218

Qy 69 -----NKKKFKPE-----YLTWEVLELKNLSLTPVEKNLHFWIGG 105

Db 219 KGIEILVKOGELLSSLINVKNQKQNSKINLLGIEYQSHKV-DIEPFIHAWVAG 277

Qy 106 QINDTAINYINQWQVNSDYTVKVFYDSNAF-----LINTLKKT----- 144

Db 278 APPDNTFSYITAFNLTKYKDYTVLLWIDNFAAFAKSGILKNIAMNYAIRMRLRRTPHILA 337

Qy 145 -----IVSATNTTLE--SFRENLDPEPDY-----NKFYRKMEIYYDKQKHF 186

Db 338 FEMNEVILKIQNIQNETTIEPKETRERLKELENRYKSLTSETKEKFNVPFLSMEIGMQDNY 397

Qy 187 IDYKQSQIEENPEPT-----IDNIK--TYLSNEYSKOLEALNKYIEBSLNKITANNNG- 238

Db 338 FTYCISNGISNTDDISRLDLFTNLVLSPEVQNDKSTVEKNKRDIDLLKNTISQKFGDR 457

Qy 239 ----DIRNLEKPADBDLVRLYNQELVERWNLAAASDIIRISMLKEDGGVYLDVDILPGIQ 294

Db 458 FQLRDINTLESFKPKQDYFFYQOEMLLRWNAASQVRINILKEYGGIYTDITDILPAYS 517

Qy 295 PDLFSINKPDSITNTSHEMIKLEAIMKYK--EYIPG--YTSKNFMDLDEEVQSFESAL 350

Db 518 DKVSQIINE-KSDDKRFPEDLKLRIISSESILSLIKGEKYSIKH-DGLDETTLNQNNIL 575

Qy 351 SSKSKSBEIFLPLDDIKVSPLEVKIAPANNSVI-----NQALISLSKDSY 394

Db 576 -SEIEK-----LTIDY-FKPVETKVVRDTFKIFKRYQKWTENTWIRGNPNFMTKSGK 629

Qy 395 CSDLVINQIKRYKILNDNLNPSINEGTFNTMTKIFSDKLASISNE-----DN 443

Db 630 CIDPILSGOKQYLEL-QRIRDNISYNNLFYTT-----EDLKSNNVAIGGIPAKKYLEH 683

Qy 444 MMFMIKITNYLKVGPAPDVRSTINLSGP----- 471

Db 684 GLF-----SEYRQDGPYVWSTNLNSGPDMMRMQMKYKSLGRIGEVHIKDNKLSDVNF 739

Qy 472 -GVYTGAYQDILMF---KDNSTNIHLEPE-----LRNFEPPTKISOLT-----EOEIT 517

Db 740 LGVYASSNKDNKSNFWLNPVSGVINDITPDDESSWAVNRNDINKILPEKINCHVPEKLP 799

Qy 518 SLWSFNQARAKSQFBEYKGVFEGALGDDNDLFAQNTVLDKDYVSKKILSSMKTRNKEY 577

Db 800 SLV-----YBIDSRSPFGW---DNKSIKHVTEINKDLI--KQINLLLTSSNID 843

Qy 578 IHYIVLOQDKISYEASCNLSPKDPYSILYQKNTEGSETAYYVYVADAEIKDYRIP 637

Db 844 VKLLIKL--DRELYAISSKI-----DNPLALRSIRTLOQLANTVTSNTPENTINFI 895

Qy 638 YQISNKRN-----IKLTFIGHGSEFNTDTPANLDVDS-----LSSEIE 676

Db 896 YDFYRKQDQLLSALK-----FSRNDATKIIVTNSVNEKRVFLREVLS 941

Qy 677 TILNLAKADISPKYIEINLLGCMNPFYSIYABETYPGKLLKIKORVSELMPISIQDST- 735

Db 942 CVLRSKVD---SYINEN-----KNLSKEDA--GALRDYAKLKMKELPFMLDDGDKY 989

Qy 736 -TVSANQYEVRIENSGKREILDHSGKWNKEBSIIKDISKEYIISFNP----- 782

Db 990 KIITTNAY---IKERDKL-----SGIYIENSIISGHESFDIIRSNQHEWGDLSVTEOF 1041

Qy 783 KENKJIVKSKYLHELSTLLQELIRNNANSDDILEKKVMLTECEINVAS-----NIDRQIV 837

Db 1042 KKFEPYVASE-LSSAKSIFDDIKNK-YITDPTKKNVLVHQLDSDIKERIAFLDISHYAY 1099

Qy 838 EGRIEAKNLTSDSINYIKNEFKLIESI-----SDSYLDLKHONG 877

Db 1100 PGSLEKQLS---GYVPSDINIABVLLASVGVSHGVVYVYVAPSDKLELLRRHT 1155

Qy 878 LDDSHIFISFEDISKTENGFRIRFINKETGNSI---PIETEKEIFSEYATHISKEISNIXD 934

Db 1156 KENSEWI---EKITP-----YVVDILSDNVSNVLRPLSEBQKILANDIKLEISKVS---E 1206

Qy 935 TIFDNNVNGKLVKKNVNLDAAEHVNTINSAFFIOSLEYNTTK-----ESLNLNSV--- 983

Db 1207 QYFMKLTBQKSSVIGKYSVDFDRYNEMLFLSLPINQNLTLPFMYRYPEMLDIHIGIIE 1266

Qy 984 --AMKVQVYQALFSTGLNTITDASKVVELV-----STAL-DETID 1020

Db 1267 NKANREFFYKPSLNLDFLINDERVNLGLIKKYKYLSEIHRITLNSFPADISIP 1326  
 Qy 1021 LLPTLSEGLPIATII-----DGVSLGAIAK-----ELSETNDPLL 1056  
 Db 1327 LLQTI---CPSITIIKTEYGHQLTNMTAVSVKPYDFNSLGAINDKSVSDVPA 1383  
 Qy 1057 ROEIE-AKIGIAVNLTAATAIVTSAL-----GIAGFSFL 1092  
 Db 1384 HTIVEQAKNLSWDFYNTHASITWTIARQHKSTNIEHPQSLLFDRSKGKGLSL 1443  
 Qy 1093 LVPLAGISAGISLVNN-----ELILODKATKVIDYFKHISLAETEGAFLLDDKIIIMPOD 1148  
 Db 1444 YLDTGGYGGYQKLRHNDTASLYQTK-----YNDNLKSLNRDDFFLRKTQRIITWSN 1497  
 Qy 1149 DLVLSIDNNNSITIGKCEIWRABEGSGHTLTDI--DHFFSSPSITVRKPLWSIYDVL 1206  
 Db 1498 EL-----GNRLKNAQLEVLK---DPILTEGILYQRRISLITTEYHSLAQGIS 1547  
 Qy 1207 NIKK-----EKIDFSKDLAVLPNAPRVFGYEMGWTTPGFRSLDNDGTLL 1251  
 Db 1548 SFRVTDNPGHCDHSLAQLTFIKNI-----TSNR-----NFSLYGSGI--- 1589  
 Qy 1252 DRIRHY-EGQFYRYFAFIADALITKLPRYEDTNVINLDG---NTRSFIVPVITTEQ 1307  
 Db 1590 --VKIYFSESLNNWKYIKLPVQTGSLRDIVLTTPKLSLSTGGSLNINHLVPV----- 1642  
 Qy 1308 IRKNLSYSP-YSGGYSLSLSPYNNVNDLNVENDTWIDVNDVNVKNITTESDEIQ----- 1363  
 Db 1643 ---SFYIDGGV-----INGNRISSEST---DVKNKIRSLKINGDILQHYI 1681  
 Qy 1364 ---KGELNLTSLKINTEDNKI-----ILN---NNTINFY 1392  
 Db 1682 NTHYLSEETQKIKOIVDFLQDNITIKVLESIDIKPISEIQOPLHSILSROKEHVKNLL 1741  
 Qy 1393 -GDINE-SNRF-----ISLFSILEDINI-----IIBI---DLVSKSYKILLSGNCMK 1435  
 Db 1742 SGLLDFSNKLRQGLSLKTLNVLNNFKESKINSITVEVTDLQGLRYRVDIDTRVIG 1801  
 Qy 1436 LI--ENSSDIQKIDHIGNGEHOQKIFYSYIDNETHYNGFDYSK-----KGLFTAEPFS 1489  
 Db 1802 LTFKEGINSLSALEHMN-----IDAIMSVIGLQVYARMKNDONISALDHA 1848  
 Qy 1490 NE-SIIRNLYMPDSNNLFIYSSKDLADIRINKGDYK--LLIGNYFKDKMKVLSLFTIEDT 1547  
 Db 1849 GAVSDIKNIV-----DKFLGGLITLTNNRYNPGVGSGASLEGFTSSGLEVCA 1897  
 Qy 1548 NTKLNGV---YLDENGVAQILKFMNNAKSALNTSLSMNFLESINIKNIFYNNDL--- 1600  
 Db 1898 ---RMGGTAGRYL--SNVAKVIK-----PLLDI--GINWSLYDSSLNHAKA 1938  
 Qy 1601 -PNIEFI---LDTNP--IIGSNSIGQFE---LICDKKNIOPIYFNFKETSITLYV- 1650  
 Db 1939 TQIEYISTAIDVFSFSINTALSIGAAYPLAIA-----IVPITI-FSHEVKNAVAVYN 1992  
 Qy 1651 --GNRQNLIVEPSYHLDL-----SGNISSTVNFQKLYG---IDRYVNVKVII 1694  
 Db 1993 QINERHKULAEKYLDSGAKVLSINKATGIIDLSNNQVLNGLNYLDMRENPPIL 2047

RESULT 7  
 T28676  
 rhoptry protein - Plasmodium yoelii (fragment)  
 C;Species: Plasmodium yoelii  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
 C;Accession: T28676; A45521  
 R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
 Mol. Biochem. Parasitol. 76, 329-332, 1996  
 A;Title: Comparison of two members of a multigene family coding for high-molecular mass  
 A;Reference number: Z20507; MUID:97077455; PMID:8920022  
 A;Accession: T28676  
 A;Status: preliminary;  
 A;Molecule type: DNA

A;Residues: 1-2401 <SIN>  
 A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41363.1  
 R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
 Mol. Biochem. Parasitol. 42, 241-246, 1990  
 A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c  
 A;Reference number: A45521; MUID:91101660; PMID:2270106  
 A;Accession: A45521  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 2260-2401 <KEE>  
 A;Cross-references: GB:M34281

Query Match 5.4%; Score 470.5; DB 2; Length 2401;  
 Best Local Similarity 19.7%; Pred No. 2.6e-09;  
 Matches 409; Conservative 328; Mismatches 671; Indels 673; Gaps 102;

Qy 27 ILNALEEYHNHSSSVVEKY-----LKLKDINNLTNDYLNLYTKKSGRNLKALKEFK 76  
 Db 62 LINKVSDGNIIEENDIDNFDLSKPNFKLLEGLNGVFNKWNKYK--NNKLDQFK 118  
 Qy 77 EYLTHEVLELKNLSLTPVEKNLHFWIGGQINDTAINYNQWKNVSDYT-VKFVYDSNA 135  
 Db 119 DTMKKIILLI-----IQYMNBEFKGLNDAMTKLKNEGISQK 153  
 Qy 136 FLINT-LKKTIVTESATNTLTESFRENLDNPFVYNKFKRMEIITYD----- 181  
 Db 154 FVINQIIOKQPKSTYDEKKEGFSSL---ELAKN-WEKKLEIITELKKNEETVOLDI 209  
 Qy 182 KQKHFIDYKQIENPEFIIDNIKTYSN-----EY-----SKLEALNKYI 225  
 Db 210 KIRELIKQKDIIEBQK--IVNDLKLKKIKRIBTEKIEYIKKAVDLKKEIKDNVYI 266  
 Qy 226 BESLN---KIT--ANNENDIRN-----LEKFAEDLVRLNQ--ELVERWNLAAASDIL 272  
 Db 267 DELAKEPPYQTYKTEKKNEIYNTIKSPDKIYVGDIEQLYNEMPSVVQESNIEHNKT 326  
 Qy 273 RISMLK-EDGGVYLDV-----IMKYKEYIPGY-----TSKNFMDLD 339  
 Db 327 ELILTKIDVYNNIQMNETETVSKHLKNIETNNKLSETILDIKIYGEITNELNKT 386  
 Qy 304 -----PDSITNTSWEMIKLEA-----IMKYKEYIPGY-----TSKNFMDLD 339  
 Db 387 EDFKNEKGLSNKIDYAKENVQNLVYKSNILEIKKHYNQDQINDIDNKEAKQNYDFK 446  
 Qy 340 BEVQSFSSALSSKSDKSEIFLPLDDI----- 366  
 Db 447 EHMKTIPPNEKMYQKPSIEIKIMKDEFLSKVKNYNDPKVYKVESEHNKFTELTNKIK 506  
 Qy 367 -KVSPLVKIAPAN-----NSVINQALISLKDSY-----C--SDLVINQI 403  
 Db 507 TEVSDEEIK-KYENKFNDSKSLINETKKSIEBEYQNIITLKKVDDYIKVCLNTNELIINC 565  
 Qy 404 KNYKILNDNLNPSINEGTFNTMTKIFSDKLASI----- 438  
 Db 566 HNKQTTLKLQNQIKTIKETNSIDKIVTDKFNILTDKKTETLTKFTGLSLNNHESNNK 625  
 Qy 439 -----SNEDNMFM-----IKITNLYKVGAPDVRSTINLSGPV 473  
 Db 626 ELLTYFYDLKANLGNKNENLYKQNEKEKAVEDIKKNV-----DINKIVSNIEITI 678  
 Qy 474 YTGAYQDLLMFKDNSTNIHLEPELRNPFEPKTKISQLTEQBITSLWSFNQARAKSQPEE 533  
 Db 679 YTSIYN---INEDTENEIG-----KSIELNTKVLKVKANVTNL-----NEIKE 720  
 Qy 534 YKGYFEGALGEDDNLDPAQNTVL--DKDYVSKKTLSSMKT-----RNKEYIHYIVQLOQ 586  
 Db 721 KLKDYDFQDFGKEKNIKYPDENKIKNDITLQKIDKSIETLTKKNS--NHIDEIKG 778  
 Qy 587 --DKISYASCNLFSPKDYSSILYQKNIEG-----SETAYYYVVADEIKEIDKIRIP 637  
 Db 779 QIDKLKVPNKTMFNEPKK-----IEKKIENIVEKIDKKKNY-----KEIDK--LL 824  
 Qy 638 YQISNKRNIKLTFF-----IGHGKSEFNTDTFANL----- 666

Db 825 NEISKIENDKTSLEKLNINLSYGS-----LGNLFQQIDEEKKAEHTIKAMEAYID 878  
 Qy 667 DVDSL---SSRETTLNL---AKADISPKYIEINLLGCNMFPSYIY-----706  
 Db 879 DLONIKKKSQIEKEMINMDIKMDI---HKEMKALINSHDDYKIYHTTSKKNHEKISDI 935  
 Qy 707 -----AEETYPGKLLKIDRVSELMPSISQDSITVSANQYEVRIINECKREIL 755  
 Db 936 RKNSLKIITQDSEESYINDIKKEKLVLE-----SQNNNT-DINQYLSKI--ENIYNIL 987  
 Qy 756 DHSGKWINKESIIKDISKEYISNPENKIIIVSKYLHELSTLLQBEIRNNANSDDIL 815  
 Db 988 K-----LNKIKIIDKV--KEYTDEIKNNK-----KINAEIS-----NSEKIITQL 1027  
 Qy 816 EKKVLTCEINVAENIDQIVGEIEEAKNLTSDSIN-----YIK-----856  
 Db 1028 KENSLKCSQKIRKTDIDNYSVCEIKNITLKIYVNEKNNINTYFKNABEYNQVSLN 1087  
 Qy 857 -NEFKLIESISDLSYDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NSIFIE 912  
 Db 1088 FNNIEMADTKSQYILNIIKNGTNNYDYNIKELKHKKSNNYK-----DEAGKNTQSIK 1142  
 Qy 913 TEKEIFSYATHI-----SKESINIKOTIFDNNVNGKLVKNLDAAEVNT-----958  
 Db 1143 KKNLEFKYEQBVTLLNKKYAVELKNKFKDKT-KNYSEQIIKEIK--DAHNTFTSQADKS 1199  
 Qy 959 ---LNSAPFIQSLIEYNTKESLNSLVAMKVQVQAQLFSTGLNWTID-----1003  
 Db 1200 EKKNEIRNEQIRIEDEAVAKNNKSKAI-LDQLSVPEPKIKPLKIKOLRTKSDCLKET 1258  
 Qy 1004 -----ASKVVELVSTALDETIDLPTLSEGLPIIATIDGVSIGAAIKELSETNDPUL 1056  
 Db 1259 KDIEIKINSLSDTQETKLIENKNINLTLEKLESKKN--QXKNIEDQKKELDEVNSKI- 1315  
 Qy 1057 ROEIEPAKIGIMAVNLTAASTAIWTSALGIASGFSILLVPLAGISAGIPSLVNNELILODK 1116  
 Db 1316 -KNIESNVNQHKNY-----EIGIVEKI-----NEITA--K 1342  
 Qy 1117 ATKVIDYKPHISLATEGAFTLLDDKIIMPQDDVLSEIDFNNSITIGKKEIWRPAGGS 1176  
 Db 1343 ANK-----DQIESQKLIIPITKILSPFKANDLEGIDITKN--LGK--YNTENNN 1389  
 Qy 1177 GH-----TLTDDIDHFSSPS---ITRKPWLSIYDVLNKKIEKIDFKOLMVLNPNRNV 1229  
 Db 1390 IYEEPIKSYDLITHLETVSKEPIYEQ-----IKNKRIITAQNELLTNIKVNKA 1439  
 Qy 1230 FQYEMGWTPGFRSLNDGDKLLDRIDHYEG-----QFYWRY-----FAPFADALIT 1276  
 Db 1440 KSY-----LDDIEANEPDRIVTHFKNKLVNDKFTWNEYSKVNKGFNINSN-IN 1488  
 Qy 1277 KLPYEDTVNRINDGNTRSFIVPVITE-----QIRKNLSYFYSYG 1320  
 Db 1489 NVK-KSTDENLLNLNLTQETKEMYANIVSKYYSKYEAENIFINIPFKLANSINIQKSS 1547  
 Qy 1321 G-----SYLSLSPY---NMNIDNLV-----ENDTWIDVNVKNITIESEIOKG 1365  
 Db 1548 GIDLFPNINIALPVLDSQKQDTLFIQSPKTSYTKISDYSYNTLLDLKRSQELQK 1607  
 Qy 1366 E-LIENILSKLNIENKIIILNHNTHINFYGDI-NESNRFISLTFISLEIDINIIEIDLVS 1423  
 Db 1608 EQQALNLIENRLLHDKVQATNELKDTLSDLNKKKEQILNKKVLLHKSNNELNKLSCNSQ 1667  
 Qy 1424 SY-KILLSGNCMKLJENSDDIOQKIDHIGFNGEHQKIYFYSYDNETKNGFI-DYSKKE 1481  
 Db 1668 NYDTTLESSKYDKIEKSNYKEKENLGIN-----FDVKAMEEQFNNDIKDIEKLE 1719  
 Qy 1482 GLFTAFESNESIIRNINYPDSNNLIYSSKOLKDIRINKGDVLLIGNYFKDKMKVSL 1541  
 Db 1720 NNY-----KHSEKONTFSENNNLIQSKKUKELT-----NAFNAEIK----1758  
 Qy 1542 FTIEDTNTIKLNGVLDENGAQILKFMWNAKSA-LNTSNSLMNPLESINIKNIFFNNLD 1600

Db 1759 -----KIEDKIIIEKNGL--INKLIETROCKMFLFYTKL-----VETLIKIKTTDY-----1800  
 Qy 1601 PNIEFILDNTFIIISGNSIGOFELICOKDKNIIQVFNPKIKETSITYLVGNRQ--NLIV 1658  
 Db 1801 -----TKFITSATFSKEFLKVIDATSNLNDN--TLQTKYDLNQNKHVASHWA 1850  
 Qy 1659 EPSYHLDDSGNI-----SSTVINFSQKYLIGIDRYNVKV 1692  
 Db 1851 DAT---NDNNNLIEKEATKTINNLT--LFTIDS--NKI 1884  
 RESULT 8  
 T28677  
 rhotry protein - Plasmodium yoelii  
 C:Species: Plasmodium yoelii  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
 C:Accession: T28677; C45521  
 R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.  
 Mol. Biochem. Parasitol. 65, 171-177, 1994  
 A:Title: A gene coding for a high molecular mass rhotry protein of Plasmodium yoelii.  
 A:Reference number: 220508; MUID:95021522; PMID:7935623  
 A:Accession: T28677  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2269 <KEE>  
 R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
 Mol. Biochem. Parasitol. 42, 241-246, 1990  
 A:Title: Identification of the gene for a Plasmodium yoelii rhotry protein. Multiple clones  
 A:Reference number: A45521; MUID:91101660; PMID:2270106  
 A:Accession: C45521  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 2131-2269 <KE2>  
 A:Cross-references: GB:M34283  
 Query Match 5.0%; Score 437; DB 2; Length 2269;  
 Best Local Similarity 20.2%; Pred. No. 3.9e-08;  
 Matches 396; Conservative 310; Mismatches 659; Indels 592; Gaps 100;  
 Qy 31 LEEYHNSSESVKYLKL-KDINNLTNYNTYKSGRNKALKKFKYLTWEV-----83  
 Db 71 LETITLTKSN--BETVRLKEIRELFKKYLD-----BEARKYLEGLKLELNKKIKD 121  
 Qy 84 -----LELKNLSLTPVEKNLHFIWIGQINDTAINYINQWQVNDYTVKVVYD 132  
 Db 122 IIAKIEYVNTVLEKKE-----IEKNYAI-----DELAN-----QSPYKVTGYIE 162  
 Qy 133 SNAFLINTLKTIVESATNTLTLESFRENLN-----DPEFDYN-----KPYRKRMEIYYDK 182  
 Db 163 NKNTIYNTI-KSYPDQIYEGDIDTFYNELSSIVKEDPIDDIETKLENLSKIDNVYDK 221  
 Qy 183 -QKHFDIYYSK---QIEN-----PEFIDNIITKYLNSYKSLBALNKYIESLKNKITAN 235  
 Db 222 IQKMEIETVKSHLNNIETNNKLPNTILE--IKKYIYDEISKE--LNKMLEDFKNK-EKE 275  
 Qy 236 NGNDIRNLEKFADEDLVRLYNQELVE---RNLLAAASDILRLSMLEKEDGVYLD--VDILP 291  
 Db 276 LSNKISDYDKREQ--LSEYKSKMLETRNNHNSQNTVNDTKBEBAKQNYKSNHEHTTIP 333  
 Qy 292 GIQDPKPSINKPDSITNTSMEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSPESALS 351  
 Db 334 TNEDEISKIIEV-----KTMKDEILSKVNTYI-----DFNKKYKETVNSEHSQPT 379  
 Qy 352 SKDSKEIFLPLDDIKVSPLEVKIAFA---NNSVINQALISLKDSY-----394  
 Db 380 ELTDKIK-----AEVSDKELKKCEQSFNDKNSLINETKNSIEKEYQNINTLKKVDEVI 432  
 Qy 395 --CSDL--VINQIKNRYKILNDNLNPSI-----NEG 421  
 Db 433 KVCKSTKESITKFSQOTILKMLNQNIKTVKETSIDKSYIEKFEQILTGKQTKLENKF 492  
 Qy 422 TDF-----NTTMKIFSDKLASIS-NEDNNMP-----446



Db 493 TEFSLNNHANNELIKYFSDLKANLGINENWLNQFTEKTEKTFNDIKEKNIHNEIS 552  
Qy 447 MIKITNLYKV-GFAPDVRGTINLSPGVVTGAYQDILLMPKONSNIHLEPELRNPF-- 503  
Db 553 KIEIKIHASIVNISETEREIGINELNKTVPKEV--KENVTNLKIKELKHDFSD 609  
Qy 504 -----PXTK-----ISLQTEQEIISLMSFNQARAKSQFEEYKGYFEGALGE 545  
Db 610 FQKEGNIKYTDKIKINDIMAVSQOIQHINGLODI-QKSESYSVSEKKE-----Q 660  
Qy 546 DNDLPAQNTVLDDKQVSKILSKMTRNKYIYHIVQLQGDKIYVEASCNLFSKDPYSS 605  
Db 661 INKLEKVSNTESNDNVE-----GIKKKQOI-IVTKIDKKNIYBEINKLLSB----- 707  
Qy 606 ILYQKNIESETAYYYYVADAEIKDKYRIPIQISNKNKNIKLTFIGHGKSEFNTDTTAN 665  
Db 708 ---ISKIEDNTS-----LEKVKOINLSYG-QNLGNLFLEQIDBEKKAE-NTIKS 753  
Qy 666 L-----DVDSL---SSEIETILNLAKADISPKYIEINLLGCNMFYSIYABETYPGKLL 717  
Db 754 MEAYIDDLNIIKKKSOEIEEMDI-KMDIN-KEME-----AL 788  
Qy 718 KIKDRVSELMPSIQDSITVSANQVEVRINEBGRKREILDHSGKWIN--KEESIIKDISK 775  
Db 789 KISHDDDKCHKDKSN-----HKNISDIYDKSKIIQDFSRSDINDIKNK 835  
Qy 776 EYISFNPKEN-----KIIIVSK-YLHELSTLLOETNNAN 809  
Db 836 LQKVSESQNHNSDINOCLNEVANIYILKLNKIKIKIDKVEYSETEKKNKINDELN 895  
Qy 810 SSDIDLEK---KVMLTECINVASNIDROIVEGREE-----AKNLTSDSINVIKNEFK-- 860  
Db 896 NSEKVIKIEGLSLKESKINSITLDDKIDECIKNIVLNKKNILNEETN-ITHFKN 954  
Qy 861 -----LIESISDLYDKHQGLDDSHFISFEDISK---TENGFIRFIN 902  
Db 955 EYVKNIVLSPNFIENMADNKSQYILEIKKNGTND-HDYNIKELSKHDKSNGYK--T 1009  
Qy 903 KETGNSIFITEKEPISEYATHI-----SKBISNIKDTTFDNNVGNLKVKNLDAH 954  
Db 1010 EADONKKAIOKQKELPEQYKEEVTLNKNKYAYAVELKNKFDKT-KNDSKQIIKEIK--DAH 1066  
Qy 955 EVNTLNSAFFIQL-----IEVNTTKESISLNSVAMQVYVAQFSTGLNTITDAS 1005  
Db 1067 NYCTLESSEKSKKNEIKNEKIHIEDEVANNKSKAIT-SIKVSVEPPKT---KIIKIN 1122  
Qy 1006 KYVELVSTALDETIDLLPTLSBGLPIIATIIDGVSLGAAIKELSETNDPLLRQ----- 1058  
Db 1123 EIRTKSDDCLEKTNDEKQIS-NLSIDTQETKLTENGKQKLTLELLESKKQKKNIEDQ 1181  
Qy 1059 -----EIEAKIGIMAVNLTAASTAIVTSALGIASGFIILLVPLAGISGIPSLVNNELIL 1113  
Db 1182 KKELDEVNSKI-----KNIENTVN----- 1200  
Qy 1114 QKATKVIDYFKHIS-LAETEGAFTLDDKIIMPQDDLVLSIDENNISI-----TLG 1165  
Db 1201 QHKKNYVEIGIVEKINEIAKTNKQNESTKELIKPTIQHIIIS--SFNANDLEGIDSDENIG 1258  
Qy 1166 KCEIWRAGEGS-----GHTLTDIDHFPSSPIYRKPMLSIYDVLNIIKKEKIDFSKD 1218  
Db 1259 K---YNTENGNIYEEPIKSYNLITNVLTVSKESITYNQ-----IQNKRIDTQKE 1305  
Qy 1219 LMVLPNAPNRVFGYEMGTGFRSLDNDGTKLLDRIRDHYEQGVYRYPFAFTADALITKL 1278  
Db 1306 L--LKNIN-----VNKAKSVLDYIK--ENEFEF-----DRIVTHF 1335  
Qy 1279 KPRYEDTVNRINLDGNTRSFIVPVTTEQIRKNLSYSFYGSG-GSYLSLSPPYNNIDLN 1337  
Db 1336 KKKLNTVNDNF-----KN-EYSKVNKGPDNISINTVKNKSTDBN 1374  
Qy 1338 LVEN-----DTWVIDVDVNVKNITTESDEIOKGELIENILSKLIENIDKNKILNHTINF 1391

Db 1375 SLLNILNQTKEMYANVNVNTYYSKYBAENIFRN--IPKLANTLNK-----IKNSSGIDL 1428  
Qy 1392 YGDINESNRFISLFSILEDINIIL-----EIDLVSXSKYKILLSGNCHKLIENTSSDI 1443  
Db 1429 SKDIKIA--ILSYLSDSKTEDTLFIPSPQKKTETTYTKIDSYSILLD-----ILKSOEL 1481  
Qy 1444 QOKIDHIGFNGEHO--KYIFYSYDINETKYNFIDYSKKEG-LFTABFSPESIIRNIYMP 1500  
Db 1482 QKK-----EQOTLKLIIF-----ENRRLYEKVOATNELRGTLSDLKYKKEKILSEV--- 1526  
Qy 1501 DSNNLFIYSKDLKDI-----RIINKG---DVKLLIGNYFKDDMKVLSLSTIEDTNT 1549  
Db 1527 ---KILLHKSNELNKLSCNFQNYDTILESSKYDQVKEKSNNYKQEKELGIDFNVTDMEE 1583  
Qy 1550 IKLVGVLDENGVAQILKFMNAKSAINTSNLNFLESINIKLIFYNNLDPNTEFILD 1609  
Db 1584 -KFN-----NDIKVIEBELNENYDSENNILOSQKQKELTNKFNKFAEIKKIDDKIIEK 1636  
Qy 1610 NFIISSGNSIGQFELICDKDKNIQIPY-----FINFKIKETSYTLVVGNRQNLIVEPSYHL 1664  
Db 1637 NDLI-----DKLIETRKNCMLFHTTTLAETLKIKITDYSKFIESATKESKEFLXYI 1687  
Qy 1665 DDSGN-ISSTVINFSQKY-LYGDIDRYV-NKVIIAPN 1697  
Db 1688 GDTNSLNDDIATLQKYLQDLHQINKYVTSKLSDATN 1723  
RESULT 9  
G70163  
Hypothetical protein BB0512 - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: G70163  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: G70163  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2166 <KLE>  
A:Cross-references: GB:AE001153; GB:AE000783; NID:G2688419; PIDN:AA066876.1; PID:G268842  
A:Experimental source: strain B31  
Query Match 4.9%; Score 424; DB 2; Length 2166;  
Best Local Similarity 19.9%; Pred. No. 1.1e-07;  
Matches 426; Conservative 338; Mismatches 691; Indels 684; Gaps 108;  
Qy 3 LVNKAQLOKQYVYFRQDEYVAILNLALEYHNMSESSVVEKYL--KLKDNINLTDNVL 60  
Db 8 LVNLFLVSIVLFV-YR-QYDKRSRALDKIKKFDVLTQVN-LEDFIEDTKTEINDLAVD-M 63  
Qy 61 NTKKSGRNKALKKEKVELTWEVLELKNLSLTPVEKNLHFIIWIGQINDTAINYNQWKD 120  
Db 64 BAYORSS-IEIITKIEE--VOQKTKNKSNDPAEVEKKIAY-----HDSMLKDLDEWTF 113  
Qy 121 VNSDYTVFVYDVSNAFLINTLKTTI-----VESATNNTLBSF-----RENLDNDFP- 166  
Db 114 KVQENIQRLVDGK--IVDKLSKTLKGFNTQIDSVESNLNSVLEKFDKANKENLESIKIA 171  
Qy 167 DYNKYRKRMIY--DKQKHIFIDYKSIQIENPEFIIDNIK-----T 208  
Db 172 SWEPFDTNKLVPKIDNLNKEISLYEKDLANIEBKNDILVKGNEKLDLBSDFLEKVE 231  
Qy 209 YLSNEYSKDLAL-----NKY--IEESLNKLTANNGDIRNLEKFADEDLRLYNQELVE 261  
Db 232 FNIGKYSKEISSESNFYENKYLKLENSIELIMESVKNKINEKEDF-----ILNRLNEEQN 287  
Qy 262 RWNLAASADIL-----RISMLK--EDGGVYLDVDI---LPGIOPDLFKSINKPDSITNT 310









Db 1729 LKTHGLEPCTTLSPSSIVETAIDHVS-----LLQAVKTKTSRMHTLIQSILGERI 1779  
Qy 1484 FTAEPFNEGSIIRNIYMPDSNNLFIYSSKDLKDIRIINKGDKVLLIGNYFKD---DMKYSL 1540  
Db 1780 F-----KLFKNSGVQSDSKI-----SIDRVQLVDESDSATIDFTVIKDKORSQKSI 1826  
Qy 1541 SFTIED-----TNTIKLNGVVLDENGVAQILKPM--NNAKSALNTNS 1581  
Db 1827 TWGIESLAGSFRKFSASMEHVITGTGLDLD-LGMTVVSILVQVRLVEAQGGKDALAVANL 1885  
Qy 1582 LAMNF-----LESINIKNI 1594  
Db 1886 VMNLKIALE-VSIGNV 1300

RESULT 12  
C16222  
hypothetical protein PF80145c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: C71622  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: C71622  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1979 <GAR>  
A:Cross-references: GB:AE001375; GB:AE001362; NID:G3845105; PIDN:AAC71819.1; PID:G384510  
A:Experimental source: clone 3D7  
A:Genetics:  
A:Gene: PF80145c

Query Match 4.6%; Score 398; DB 2; Length 1979;  
Best Local Similarity 19.7%; Pred. No. 8.3e-07;  
Matches 373; Conservative 287; Mismatches 605; Indels 626; Gaps 88;  
Qy 53 NNUTDNLNTYKSGRANKALKPKFKEYLTWVEVLELKNSTPVEKNJHFTWIGQINDTAI 112  
Db 93 NNNDDNN 149  
Qy 113 NYINQWKNVNSDYTVKVVYDSNA---FLINTLTKTIVESATNTNTLESFREN---LNDPEF 166  
Db 150 NYESKIEELEKE--LEKVDKNIDNNNDYENKLEK--EDFVKQIDMLNEKENLQEKEL 205  
Qy 167 DYNKPYRK---RMEIYYDKQHFIDYKQSIENPE-----FIIDNIITKYL 211  
Db 206 DINKREKKINEKEKNIKKETPHNTEKEYLEKKNKERETISIEID--IKKHLEKLIKBI 263  
Qy 212 NEYSKDLBALN-----YIESLNKITANNNDIRNLEKFADEDLVRLYNQ 257  
Db 264 KKKEDLENLKLLSKENLVKELKGVKFKNETINSNDNIIEKKY-----KLLEY 317  
Qy 258 ELVERNWLAASDILRAISMLKEDGGVLDVDPGLQPDLEKSIKPDISTWSEMILK 317  
Db 318 ELEEK-----NKQIDL-----LNQKEKEKEKEKEKE 345  
Qy 318 EAIMKYKEYIPGYSKDFMLDEEVSFESALSSKSDKSEIFPLPDDIKVSPLEVKIAP 377  
Db 346 KEREKE-----KEYDTLIKEL-----KDEKISILEKVHSIKVEMDIEKRE 388  
Qy 378 ANNSVINQALISKYSQSDLVNQINRYKILNDNLNPSINSGTDFNTMTKIFSDKLAS 437  
Db 389 HNFLHMEQDLKLGNSFVKV--NNQLK-VYKBEIKNLK-----TLEKKEKELKD-IEN 438  
Qy 438 ISNEDNMFMFKITNYLVKVGFAPDVRSTINLSGPGVYTGAYQDLMFKDNSTNIHLLEPE 497  
Db 439 VSKEE-----INKUINQLN-----EKEKQILAFNKNHKEBIHGLKEE 475  
Qy 498 LRNFEPFKTKISQLTQETISLWSPNQ-----ARAKSQFEYK 536  
Db 1126-1132, 1998

Db 476 LKE-----SVKITIETQELQEMVDIKOKELDQLOEKYNAQIESISIELSKEKEYNQYN 531  
Qy 537 GYFEGALGEDNLOFAQNTVLDKDVSKILLSSMKTRNKEYIHYIVQLQGDKISYEASCN 596  
Db 532 TYIE-----EINNLN-----EKUEETNKEY-----TNLQN---NTNIN 563  
Qy 597 LFSKDPYSSILYQKNIEGSETAYVYVADAEIKIDKYRIPYQISNKENIKLITFIGHGKS 656  
Db 564 MLNND-----IHLNGNIKTMT-----QISTLKN----- 588  
Qy 657 EFNDDTFANLVDLSSEIETILNAKADISPKYIEINLGNMPSYSIYAEETYPGKLL 716  
Db 589 -----DVHLINQIDK-LNNEKGLNSKISELNV-----QI 618  
Qy 717 LKIDRVSELMPSISQDSITVSANQYEV--RINEGKREILDHSGKWINKESIIKDISS 774  
Db 619 MDLKEEKDFL-----NNQIVLSNQIDILTRMBEKENKMLEQENKYKQEMELLRGNLKS 673  
Qy 775 KEYISFNPKENKIIIVKSKYLHELSTLQEIIRNNANSSDIDLEKKVMLTECEIN-VASNID 833  
Db 674 SENILANDEE-----VCDLKRKLSLASEMOMKEED 706  
Qy 834 QOIVE-----GRIEAKNLSDSINYIKNEFKLIESDSLYDLKHONGLLDSSHIFSE 887  
Db 707 KKLAEKDDCDVRIREMNEKNEKDKINMLKBEYE-----DKINTLKEQN-----E 750  
Qy 888 DISKTENGFRIRFINKETGNSIFITEKEIFSEYATHISKEISNIKDTIFDNVNGKLVK 947  
Db 751 D-----KINTLKEQ--NEDKINTLKE-----EY-----EHKINTMKEVEHKN-----T 788  
Qy 948 VNLOAAHEVNTLNSAFFTQSLIEYNTTKESLSNLSVAM-----KVQVYAQLFSTGLNTI 1001  
Db 789 LNEQNEHKINTLNE-----QNEHKINTMKEEYEDKNTLNEQNEQDNLSKBEYENKINQI 844  
Qy 1002 TDAS--KVVELVSTALDETIDLLPFLSEGLPIIATIIDVSLGAAIKELSETNDPLRQE 1059  
Db 845 NSNBEIKIKDVVNEYIEEVKLVLDKDKKQFDEKINVAHIKAHEK----- 892  
Qy 1060 IEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGISPLVNNELLQDKATK 1119  
Db 893 -----QILLTEM-----BELKQ--RONK 909  
Qy 1120 VID-YFKHISLAETGAFTLLDDKIIMPQDDLVLSEIDFNNNNSITLGCETWRAB--GGS 1176  
Db 910 YSDLYEKYIKLAKSICMIINIECCDDIENEDIIRIEEYINNNGKLGK-EVEEKEHGRHS 968  
Qy 1177 GHTLTDDIDHFPSSPSITYRKPWLSIYD--VLNKKKEKIDFSKOLMVLNPNRNVFGE 1233  
Db 969 SPNILKSKEKFPKN-----SIEDKSHELKKKHEKDLKSKDEI--BEKNK----- 1011  
Qy 1234 MGWTFGPFSLNDGTKLADRIR-----DHVEGQFYWRVYFAFIADALITKLKPRY 1282  
Db 1012 -----KIKELNNDIKKLQDEILVYKQNAQOVDHKKKSWI-----LKQKSKEKI 1057  
Qy 1283 EDTNVRINLDGNTRFSI-----VPVITTEQIR-KNLSYFYGSGGYSLSLSPYNNMID 1335  
Db 1058 KDKENQINVEKNEEKDLKKODEIRLNEELVKYKTI-----LYNLKDPQLQND 1108  
Qy 1336 -LNLVENTWJD-----VDNVVNNTIETSEIQKG-----ELIENILS-----KUNTED 1379  
Db 1109 LLSKIDINSLATTNEGVCVDKIEEHILDYDEEINKRSRNLFLQKNEICSLTTEVMEINNK 1168  
Qy 1380 NKIILNHTINFGY-----DINESNRFLSLTFSILEDNIIIE-----I 1418  
Db 1169 NELISENNKLNLDVGOKKKLKKDVEKQKKEKLNQKQKCNQIDELNEEVEKLNENI 1228  
Qy 1419 DLVSKS-----YKILLSGNCWKLIENSDDIQO-----KIDH----- 1449  
Db 1229 ELIITYSNDLNKFKDMKENLMMKLEDNEDNKKMKSIDDMEKEIKYREDEKKNLNEIN 1288  
Qy 1450 -----IGFNGEHOKY-----IFVSYDNTETKNGF-IDYSKK----- 1480  
Db 1289 NLKKKNDMCIKYNEMNIKYGDICVKYBEMSUTYKETSLLKQBQIKVKYDEKCSQYDEIRF 1348

```

Qy 1481 ---EGLTAFSNE--SIIRNIYMPDSNNUIFYISKDKLKOIRIINKGDVKLLIGNYFK-- 1533
      |         |         |         |         |         |         |         |
Db 1349 QYDEKCFQYDEINKKYGALLNI---NITNKWVDSKVDRNNNEIISVDNKVEGIANYLKQI 1405
      |         |         |         |         |         |         |         |
Qy 1534 -----DDMKVSLSF--IEDTNTIKLVGVLDENGVAQILKFM-----NNAKS 1574
      |         |         |         |         |         |         |         |
Db 1406 PELNEEIIRUKGEINKISLVSENLKNSYDINKMKHQSQ-----LLPLETKNKENEK 1460
      |         |         |         |         |         |         |         |
Qy 1575 ALANTSLSLN-----FLESINI-KNIFYNNLDPN---IEPILDTNFIISG 1615
      |         |         |         |         |         |         |         |
Db 1461 IINLTQSYSDAYKKKSDSKLCAQGFVDDVNIYGNISNNNIRTNKYKVEEMFDNIEKN 1520
      |         |         |         |         |         |         |         |
Qy 1616 SNSIGQFELICDDK--NIQPYFNFKIKET-----SVTLVGNRQNLIVE---PSYH 1663
      |         |         |         |         |         |         |         |
Db 1521 GMLSKYIHLLEENKFCMKIIVENENIKSNKTIIGLYNYSRYVGLREDLCKEIVPS-- 1578
      |         |         |         |         |         |         |         |
Qy 1664 LDDSGNISSTVINFQKVLVYCIDRYVNVKVII 1694
      |         |         |         |         |         |         |         |
Db 1579 --KIGNISNKNENNKKKN-NTCDGYDEKVTI 1606

RESULT 13
T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z18935
A/Accession: T18427
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-3724 <LAW>
A/Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C/Genetics:
A/Introns: 307/1; 1545/2
A/Note: C0335C

```

[illegible]

Db 1853 IENLSEHLRHKYDIEMIPIDIVGNIR--LYKLFYINMIKNN--YSYLKQNNFF-----PY 1903  
Qy 1331 NNMIDLANVENDTWIDV-----DNVKNITIESDEIQGELIENILSKNIENDKIIL 1384  
Db 1904 KYKGRGKAKGMLMINIKRHEYNLMCK--KNGNEKTG---NNKKKKNNKEDKNDYI 1957  
Qy 1385 NNHTINFYGDINESRNFISLTFSLIEDINIIIEIDLVSYSKYL---LSGNCMKLIENS 1441  
Db 1958 NNNNNNNNNNDNN-----INV-----GYKRMQNNINPNNLYEIKNN 1996  
Qy 1442 DIQOKIDHIFNGEHO-----KYIPYSYIDNETKYNIDYKSGEGLFTAFESNESIR 1495  
Db 1997 -----FNLYHRYSLFYKYCPVSYIND-----NKIEGKKK---PCVIYKKN--- 2035  
Qy 1496 NIYMPDSNNLFI-----YSKOLKDI-----R.IKNGDVKLL----- 1527  
Db 2036 -IYMLNNHDMYIEFLKHPKPYASKVKESVCSLVIYFPNMDIQKKIYLLTYEKLHI 2094  
Qy 1528 -IGNYF-----KDDMKVS-----LSFTIEDTNTIKLN----- 1553  
Db 2095 DIGKYPKQYFFLSHQDKEQHKNNKMSNNDDVDNKIEDAHIFEHRIIEGKTISPLFIV 2154  
Qy 1554 -----GVYLDENGVAQILKFMNNAK-SALNTSNTSLMNFLE----- 1587  
Db 2155 RVICIHMKKIHOPIYIIEIITSIKLKHNFNLLIMKKKILNQCEIKKKKKKKKK 2214  
Qy 1588 -----SINTKNIFYNNLDNIEF-----ILDTFIISGNSISIQFELICDKDKNIQPYFIN 1638  
Db 2215 KKKKKKIPEKKLVNNEKKNLFLFSSIMKIKFL-----YNLIIQKQKIKR--VS 2263  
Qy 1639 FKIKETSITLYVGNRQNLIVEPSYHL-----DPSGNISSTVFNFSQKLYGIDRYNKKV 1692  
Db 2264 FK-----NVLDKLSHTNIKDKWAKDINLLVLIEDMFPFEKKYLNKV 2306

RESULT 14  
T18501  
hypothetical protein C0760c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T18501  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: Z18935  
A:Accession: T18501  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3394 <LAW>  
A:Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1  
C:Genetics:  
A:Map position: 3  
A:Note: C0760c

Query Match 4.5%; Score 386.5; DB 2; Length 3394;  
Best Local Similarity 19.0%; Pred. No. 4.4e-06;  
Matches 401; Conservative 319; Mismatches 699; Indels 687; Gaps 99;  
Qy 6 KAOLQKVVYKFRIODEYVAIINALEEVH---NMSESV-----VEKYLKLDINNLTLD- 57  
Db 699 KIYLEKLLNEKLYIKD-----IEKKLDLHGVIKNNKEDIYILQVEKQTLIKVISSVYDY 754  
Qy 58 -----NYL-----NTYKSGRNKALKFKFYLTMVELELKNLSLTPVEKNLHFIWIGGOI 107  
Db 755 TKMESENHIFKQMTTWKMLNNVHMSNKKDYNNQNNQNNQNNQNNQNNQNNQNNQNNQNI 814  
Qy 108 NDTAINYINQWKD-----VNSDYTVKPVYDSNAFLINTLTKTIV--- 146  
Db 815 ENNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNN 874  
Qy 147 -----ESATNNTLESFRENLDNPEFDYNNKFRKRMWIIYDKQ 183  
Db 875 KDEKKKTSNIKESISLKKKHTKKSITNNNDN--NNEDDDNDMLSVMSYNDVDVIKNR 932

Qy 184 KHFIDYKSOLEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEBSLNKITTANNNGDIRNL 243  
Db 933 K-----ENKEILEEHVSFSSFN-----NEYIAHSFISILLQLSNYIFNI 973  
Qy 244 EKFADEDLVRYNOELVERMNLAAASDILRISM-----LKEDGGVYLDVLDILPGIQPDLF 298  
Db 974 E-----CKOMEYFKSNLLSYVDYTTITIELFRLKKNYNIPIFIEQILGTQYPSIL 1024  
Qy 299 KSINKPDSITWTSWEMIKLEAIMKVKEYIPCYTSKNFMDLDEEVQSPESALSXSDD-- 355  
Db 1025 OKLHDGIYCLKDNKKNNKNGDNGKQEDDDGKKKNDG--DNKSQEDDGGKKKNDGDN 1083  
Qy 356 KSEIFLPDDIKVSPLEVKTAFANNSVINQALISLKDSYCSDLVINQIKRYKIINDNLN 415  
Db 1084 KSQ-----EDD-----YGNKKKNDD--DDDSYKIELIVDELNCKKNYTDDEL 1126  
Qy 416 PSINEGDTFNTMK---IPSDKLASINEDNMFMKITN---YLVK-----GP 458  
Db 1127 YELMKGSDFDIFKKYKQNFYLNHFINNIPSTIISFHIHNTEDKYKYVRYFNLFNFNF 1186  
Qy 459 APDVRSTINLSGPV-----YTGAYQDLL-----MFKDNSTN-----IHLPE 497  
Db 1187 -----SNVELSFDLIRRFDKILRLTKYEQLEENYEKIKNKNEKEYLHACIKELEMN 1241  
Qy 498 LRNPEFPKTKISQLTQEITSLWSFNQAR--AKSQFBEYK-KGYFEGALGDDNDL--- 550  
Db 1242 LERYNNEKIIIDEQINEKEKINIINEKYLILEXEYEQKNIFINA--QIENLEKEKK 1299  
Qy 551 FAQNTVLDKDYVSKKI-----LSMKTNRNKEYIHYIVOLQGDKISYEASCNLFSKDPYS 604  
Db 1300 QLQBEIIQKDMINVKLNEKNDIIKIYEKEQYLLTLQENKDSHNY-----LKDKFE 1352  
Qy 605 SILQKQNEGSETAYVYVADAETKEIDKYRIPQOISNKRNIKLTFFIGHGKSEFNTDFA 664  
Db 1353 NLL---NL--NEKUKYDH--DISLNKINTLMELEKNNKK--TF-----HMNLRVE 1396  
Qy 665 NLDVDSLSSEIETIINLAKADISPKYIEINLLGCMFYSYI-----YAE 708  
Db 1397 NNNLLKKMELQKNYNIKKELNERIKQINVRNNVSTLSLRDNSTRGRSTHQINNMYN 1456  
Qy 709 ETPYFKL-LLKIKORVSEL-----MPSIQDSITVSANQYEVRIE-- 749  
Db 1457 NTHLFGMGASKINNINISNLYSNIMHSHRGSIIKNKEDASGNSIQARMNKKDSTNIN 1516  
Qy 750 -----GKREILDHSGKWKEE 766  
Db 1517 NIHTDNNNNNNNNNTLNSINSHLYPPFPFNNVNSPKVMGCMCVTLASG--VNKK 1574  
Qy 767 SIIDISSKYSIFNPKENKIIVKSKYLHELSTLLQEIIRNNANSDDIDL--EKKVMLTEC 824  
Db 1575 DFLNLNEENESP--LEVEIRKS-----LQBELCDKESILKIKGKNILITCI 1623  
Qy 825 EI-----NVAENIDR--QIVEGRIEAKNL-----TSDSINYI----- 855  
Db 1624 ETWKCFCKNSKEEISRLKEICKEQLEKHEFLINKSNEDKLYINSLLCDEKDYIIVV 1683  
Qy 856 -----KNEF-KLIESISDSLYD--LKHONGLDDSHFISFEDISKTENGRIFRINKE 904  
Db 1684 KDIKNNMNEIDKNNNDINEKSYETKLKHEH-----NNLINEMNILK-----NKE 1729  
Qy 905 TGNSTIFETEKEIFSEYATHISKEISNICKDTIFONVNGKLVKKNLDAHAHVNTLNSAFF 964  
Db 1730 TENNMKQKEE-----DYIKLIKOKTNIQNEVD-----LLEKYN-----EYVVKNNMLY 1775  
Qy 965 IQSLIEYNTTKESL-----SNLSVAMKVQVYA-QLPSTGLNTITDASKVVELVSTALDETI 1019  
Db 1776 NDMVLLKHEKEEIEPLKENIKILOKONTYNDMFKNQINYVDN-----NLLKNRLDQUL 1830  
Qy 1020 DLLPTLSGLPIIATIIDGVSILGAAIKELSETNDPLLRQ-----EIEAKIGIMAVNLTA 1073  
Db 1831 NNQLOQLKHL-----DTNQKHEQLKQYDYIEIKERLKIETKINK 1870  
Qy 1074 ASTAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETE 1133



Db 922 NTSEHEKDER-----DGKYLEASINDYMSDDKKKRYVDSIESLRGSKI-- 965  
Qy 864 SISDSLYDLKHONGL-----DDSHFISPDISKTEGFRIRPINKETGNSIPIETEK 915  
Db 966 -KNDQIYQGGHSSLLYYDDNNDDNNNNMNDSSSSNNHYYILTNDKELNNDNFNNNL 1024  
Qy 916 EIFSEYATHISKE--ISNIKOT-----IFONVNGKLVKYNLDAAEHVNTLNSAFFIQ 966  
Db 1025 EINNQNKVIEKNLEYINNKKLTGTSNYEQSNNTNSKDEHNISSDKSKEDTLNLS--RK 1082  
Qy 967 SLIEYNT-TKESLSNLSVAMKVQVVAQLFSTGLNITTDASKVVELVSTALDETIIDLPTL 1025  
Db 1083 SSVYNNKILQSTNSKSLNGAYE--NNLFS-----GKKKKNKGTVLKD-IEHINDI 1130  
Qy 1026 SEGFLPIATIIDGVSLGAAIKELSETNDP-LLRQETEAIGIMAVNLTAASTAIVTSALG 1084  
Db 1131 QOKYP-----EDLNINCVNKVVIENEERKHLPLELE-----YNLVSSEKFP----- 1171  
Qy 1085 IASGPSILVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGAFLLDDK-- 1142  
Db 1172 -----GLNKKNNDNII-----YMKH-----QNYHNLYDDNQK 1199  
Qy 1143 --IMPQDDLVLSEIDFNNSITLGKEIWRAGSGGHTLTDIDHFFSPSPITYRKPMWL 1200  
Db 1200 KHLIPDTNKNV--SIQRNNNINSVIKTNHVEVEKNNKQORNYD-----NFTCDKXKK 1249  
Qy 1201 SYDVLNIIKKEIDFSKDLMLVLPNAPRVFGYEMGWT-----PGFRSLNDGDK-LDDRIR 1255  
Db 1250 IYNNIIN--SDKDIYHNNIITYKNEKEGIGNIHLNRNDKIDTNFELLKLDGVKEFLDTFK 1307  
Qy 1256 DHYEGQFYWKYPAFIADALITKLKPRYEDTNVRINLDGNTRSFIVPVTTEQIRKNLSYS 1315  
Db 1308 DSY-----IDCHNKKENI----- 1320  
Qy 1316 FYGSGGSYSLSPYNNIDNLNLVENDTWIDVNVVKN-----ITTESDEIQGELIENI 1371  
Db 1321 -----LNMTNKN-----KEDHQIIDVADKI FNETNMTWNNKIYDDK----- 1358  
Qy 1372 LSKLNTEDNKIILNNTINPYGDI-----NESNRFISLTFSILEDINIIIEIDLVSXS 1424  
Db 1359 ---NVHEKK-CTHNDVIHNNMDILSTSIKNNEENLF-----IDTYQKQ 1397  
Qy 1425 YKI-LLSGNCMKLIENSDDIQKIDHIGFNGEHQKVIYFYSIDNETKYNG----- 1473  
Db 1398 NRIGDIYMRINILQDDDDDDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1457  
Qy 1474 FIDYSKKEGLTAEFSESNE--IIRNIYMPDSNNLFIYSSKDKOIRIINKGDKVLLIGN 1530  
Db 1458 FSDFIEKKKIKIKKNKESVHKIDESLLSNEKN-----KVSLLIN 1499  
Qy 1531 YPKDDMKVLSPTIETDNTIKLVGYLDENGAQILKFNNNAKSALNTSNLSMNFLESIN 1590  
Db 1500 NKKDSSSVD-----NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1549  
Qy 1591 IKNIIFYNNLDPNIEFTLDNFI-----ISGSNISQFELICDKDKNIQPYFINFKIKE 1643  
Db 1550 DNN---NNNNNDSFSKNNLNNDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1606  
Qy 1644 TSYTLVXGNRONL-----IVEP-----SYHLDDSGNISSTVINFSQKLYGIDR 1687  
Db 1607 VSVSPLINHPNLSMRKORTIEPLKINGKNKLIKDLAKKIQEQVERKIRKYIQMDQ 1664



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: November 5, 2003, 19:27:13 ; Search time 13.875 Seconds  
(without alignments)  
7069.698 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1020  
Perfect score: 5192  
Sequence: 1 MNLVNAQLQKVVYKFRIQ.....ITDASKVVELVSTALDETID 1020.

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5171	99.6	2364	2 I40884	cytotoxin L - Clos
2	4086	78.7	2366	2 S10317	toxin B - Clostrid
3	4073.5	78.5	2367	2 S70172	toxin B - Clostrid
4	2783	53.6	2710	2 A37052	toxin A - Clostrid
5	1424.5	27.4	2178	2 S55805	alpha-toxin - Clos
6	408.5	7.9	3169	2 T00296	toxin B - Escheric
7	362	7.0	3225	2 D81702	adherence factor T
8	350	6.7	2401	2 T28676	rhostry protein -
9	345.5	6.7	1127	2 T28317	ORF MSV156 hypothe
10	330	6.4	3724	2 T18427	hypothetical prote
11	318.5	6.1	2166	2 G70163	hypothetical prote
12	316.5	6.1	3255	2 G81702	adherence factor T
13	315.5	6.1	3394	2 T18501	hypothetical prote
14	311	6.0	1302	2 JC6009	surface-located me
15	310	6.0	2269	2 T28677	rhostry protein -
16	310	6.0	3335	2 H81702	adherence factor T
17	302	5.8	2819	2 A90551	conserved hypothet
18	293.5	5.7	1252	2 B42771	reticulocyte-bindin
19	288.5	5.6	1447	2 F82909	hypothetical prote
20	286.5	5.5	1979	2 C71622	hypothetical prote
21	285.5	5.5	4688	2 F82885	hypothetical prote
22	285	5.5	1306	2 T28313	ORF MSV152 probabl
23	280	5.4	1516	2 E71619	RAD2 endonuclease
24	275	5.3	1939	2 T18372	repeat organellar
25	273.5	5.3	2829	2 A42771	reticulocyte-bindin
26	272	5.2	980	2 E71606	hypothetical prote
27	271	5.2	1711	2 T18429	hypothetical prote
28	266.5	5.1	1121	2 F71613	hypothetical prote
29	261.5	5.0	1183	2 F90559	conserved hypothet

vsaa-like (mycopla  
hypothetical prote  
progenitor toxin n  
conserved hypothet  
myosin-like protei  
hypothetical prote  
limpi protein - Myc  
hypothetical prote  
serine/threonine-s  
hypothetical prote  
probable major sur  
major merozoite su  
multiple banded an  
hypothetical prote  
hypothetical prote

30 261 5.0 1017 2 D90550  
31 258.5 5.0 1712 2 C71618  
32 257.5 5.0 1162 2 A47708  
33 257 4.9 963 2 C90535  
34 256.5 4.9 1875 2 S38173  
35 256 4.9 1090 2 H86806  
36 255 4.9 1365 2 T30822  
37 255 4.9 1817 2 D71606  
38 255 4.9 2485 1 H71621  
39 254 4.9 2136 2 A05037  
40 250 4.8 1640 2 A24594  
41 249 4.8 1639 2 S05603  
42 248.5 4.8 1631 1 SAZQK1  
43 247.5 4.8 1103 2 H82884  
44 247 4.8 1679 2 S48385  
45 246 4.7 1619 2 T18499

ALIGNMENTS

cytotoxin L - Clostridium sordellii  
C:Species: Clostridium sordellii  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 15-Oct-1999  
C:Accession: I40884  
R:Green, G.A.; Schue, V.; Monteil, H.  
Gene 161, 57-61, 1995  
A:Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium s  
A:Reference number: I40884; MUID:95369733; PMID:7642137  
A:Accession: I40884  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2364 <RES>  
A:Cross-references: EMBL:X82638; NID:gi1000694; PIDN:CAA57959.1; PID:gi1000695  
C:Superfamily: cpl repeat homology  
C:Keywords: cytotoxin

Query Match 99.6%; Score 5171; DB 2; Length 2364;  
Best Local Similarity 99.7%; Pred. No. 9.2e-187;  
Matches 1017; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MNLVNAQLQKVVYKFRIQEYVAIINALNEEYHNMSSESVVEKYKLDKLNNTDNYL 60  
Db 1 MNLVNAQLQKVVYKFRIQEYVAIINALNEEYHNMSSESVVEKYKLDKLNNTDNYL 60  
Qy 61 NTYKSGRNKALKKPKYLTMEVLEKNNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120  
Db 61 NTYKSGRNKALKKPKYLTMEVLEKNNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120  
Qy 121 VNSDYTVKVDNSNAFLNTLTKTIVESATNTLTLESFRENLDPEFDYKFKRMEIY 180  
Db 121 VNSDYTVKVDNSNAFLNTLTKTIVESATNTLTLESFRENLDPEFDYKFKRMEIY 180  
Qy 181 DKQKHFIDYKQIENPEFIIDNIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI 240  
Db 181 DKQKHFIDYKQIENPEFIIDNIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI 240  
Qy 241 RNLEKFADEDLVRLNQELVERWNLAASDILRI SMLKEDGGVYLDVILPGIQDLPFKS 300  
Db 241 RNLEKFADEDLVRLNQELVERWNLAASDILRI SMLKEDGGVYLDVILPGIQDLPFKS 300  
Qy 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQRFESALSSESKSKSEIF 360  
Db 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQRFESALSSESKSKSEIF 360  
Qy 361 LPDDIKVSPLEVKTAFAFANNVINQALISLKDYSVDLVINOIKRYKLTNDNLNPSINE 420  
Db 361 LPDDIKVSPLEVKTAFAFANNVINQALISLKDYSVDLVINOIKRYKLTNDNLNPSINE 420  
Qy 421 GTDFNTTWKIFSDKILASINEDNMFMKITNYLVKGFAPDVRSTINLSGPGVYTGAYOD 480

Db 421 GTDFNTMTKIFSDKLASIGNEDNMFMKIITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480  
Qy 481 LLMFKDSTNTHLLEPELNFEPKTKISQLTQEITSLWSNOARAKSQPEYKKGVEF 540  
Db 481 LLMFKDSTNTHLLEPELNFEPKTKISQLTQEITSLWSNOARAKSQPEYKKGVEF 540  
Qy 541 GALGEDNDLFAQNTVLDKDYVSKSLSSMKTRNKEYIHYIYVQLQDGIYSASCNLF 600  
Db 541 GALGEDNDLFAQNTVLDKDYVSKSLSSMKTRNKEYIHYIYVQLQDGIYSASCNLF 600  
Qy 601 DPYSSILYQKNTGSETAYYYYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660  
Db 601 DPYSSILYQKNTGSETAYYYYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660  
Qy 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720  
Db 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720  
Qy 721 DRVSELMPSISQDSITVSANQYEVRIINEGKREILDHSGKWKNEBSIIKDISKEYISF 780  
Db 721 DRVSELMPSISQDSITVSANQYEVRIINEGKREILDHSGKWKNEBSIIKDISKEYISF 780  
Qy 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840  
Db 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840  
Qy 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFISEFIDSKTENGFRIRP 900  
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFISEFIDSKTENGFRIRP 900  
Qy 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960  
Db 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960  
Qy 961 SAFFIQSLIEYNTTKESLNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020  
Db 961 SAFFIQSLIEYNTTKESLNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 2  
S10317  
toxin B - Clostridium difficile  
C:Species: Clostridium difficile  
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_change 15-Oct-1999  
C:Accession: S10317; S21894; S22434  
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.  
Nucleic Acids Res. 18, 4004, 1990  
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.  
A:Reference number: S10317; PMID:90326540; PMID:2374729  
A:Accession: S10317  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2366 <BAR>  
A:Cross-references: EMBL:X53138; NID:g40442; PIDN:CAA3298.1; PID:g40443  
R:Michel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.  
submitted to the EMBL data Library, July 1991  
A:Description: Comparative analysis of Clostridium difficile toxins A and B.  
A:Reference number: S21894  
A:Accession: S21894  
A:Molecule type: DNA  
A:Residues: 1271-2366 <BIC>  
A:Cross-references: EMBL:X60984; NID:g40445; PIDN:CAA43299.1; PID:g40446  
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.  
Mol. Gen. Genet. 233, 260-268, 1992  
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.  
A:Reference number: S22434; PMID:92293124; PMID:1603068  
A:Accession: S22434  
A:Molecule type: DNA  
A:Residues: 1791-2366 <VON>  
A:Cross-references: EMBL:X60984  
C:Genetics:  
A:Gene: toxB  
C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin  
Query Match 78.7%; Score 4086; DB 2; Length 2366;  
Best Local Similarity 77.5%; Pred. No. 5.4e-14;  
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;  
Qy 1 MNLVNAQLQKVVYKFRIOEDYVAIILNALLEEYHNMSSESVVEKYKLKIDINNLTDNYL 60  
Db 1 MSLVNRKQLEKXANVFRFTQEDYVAIILDALEEYHNMSENTVVEKYKLKIDINSLTDIYI 60  
Qy 61 NTKYKSGRNKALKKPEYLTWELVLEKNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120  
Db 61 DTKYKSGRNKALKKPEYLTWELVLEKNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120  
Qy 121 VNSDYTVKVFVDSNAFLNTLTKTIVESATNTNLTESFRENLDNDFYKFRKMEIYY 180  
Db 121 VNSDYTVKVFVDSNAFLNTLTKTIVESATNTNLTESFRENLDNDFYKFRKMEIYY 180  
Qy 181 DKQKHFIIDYKSOIBENPEFIIDNIITKYLSEYKDEALNKYTEESLNKITANNNGDI 240  
Db 181 DKQKFNINYKAQRENPPELLIIDDIVKYLSEYKDEALNKYTEESLNKITANNNGDI 240  
Qy 241 RNLEKFADEDLVRLYNQBELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300  
Db 241 RNLEKFADEDLVRLYNQBELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300  
Qy 301 INKPOSITNTSWEMIKLEAIKMYKEYIPGYTSKNFDMLEDEEVORSEFALSSEKSKSEIF 360  
Db 301 IETKPSVTVDFWMTKLEAIKMYKEYIPGYTSKNFDMLEDEEVORSEFALSSEKSKSEIF 360  
Qy 361 LPLDIDIKVSPLEKVIKAFANNSVINQALISLSDSYCSNLIQNKRYKILNDNLNPSINE 420  
Db 361 SSLGDMELASPLEKVIKAFANNSVINQALISLSDSYCSNLIQNKRYKILNDNLNPSINE 420  
Qy 421 GTDFNTMTKIFSDKLASIGNEDNMFMKIITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480  
Db 421 GTDFNTMTKIFSDKLASIGNEDNMFMKIITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480  
Qy 481 LLMFKDSTNTHLLEPELNFEPKTKISQLTQEITSLWSNOARAKSQPEYKKGVEF 540  
Db 481 LLMFKDSTNTHLLEPELNFEPKTKISQLTQEITSLWSNOARAKSQPEYKKGVEF 540  
Qy 541 GALGEDNDLFAQNTVLDKDYVSKSLSSMKTRNKEYIHYIYVQLQDGIYSASCNLF 600  
Db 541 GALGEDNDLFAQNTVLDKDYVSKSLSSMKTRNKEYIHYIYVQLQDGIYSASCNLF 600  
Qy 601 DPYSSILYQKNTGSETAYYYYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660  
Db 601 DPYSSILYQKNTGSETAYYYYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660  
Qy 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720  
Db 661 DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYARETYPGKLLKTK 720  
Qy 721 DRVSELMPSISQDSITVSANQYEVRIINEGKREILDHSGKWKNEBSIIKDISKEYISF 780  
Db 721 DRVSELMPSISQDSITVSANQYEVRIINEGKREILDHSGKWKNEBSIIKDISKEYISF 780  
Qy 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840  
Db 781 NPENKIIIVKSKYLHELSTLLQBIIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840  
Qy 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFISEFIDSKTENGFRIRP 900  
Db 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHFISEFIDSKTENGFRIRP 900  
Qy 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960  
Db 901 INKETGNSFIETEKIEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEHVNTLN 960  
Qy 961 SAFFIQSLIEYNTTKESLNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020  
Db 961 SAFFIQSLIEYNTTKESLNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 3

S70172 toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 15-Oct-1999

C:Accession: S70172; S44271

R: von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, B.; Sartingen, S. Mol. Microbiol. 17, 313-321, 1995

A:Title: Cloning in on the toxic domain through analysis of a variant Clostridium difficile

A:Reference number: S70172; PMID:7494480

A:Accession: S70172

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-2367 <VON>

A:Cross-references: EMBL:223277; NID:g761713; PIDN:CAA80815.1; PID:g761714

A:Experimental source: isolate 1470

R: Sartingen, S.; von Eichel-Streiber, C.

A:Submitted to the EMBL Data Library, July 1993

A:Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.

A:Reference number: S44271

A:Accession: S44271

A:Molecule type: DNA

A:Residues: 1-1323, 'N', 1325-2367 <SAR>

A:Cross-references: EMBL:223277

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 78.5%; Score 4073.5; DB 2; Length 2367;  
Best Local Similarity 76.9%; Pred. No. 1.6e-145;  
Matches 785; Conservative 113; Mismatches 122; Indels 1; Gaps 1

1	MYLVNKAQIQXWVYKRIQBEDEVATLALAEYHNHNESSVVEKYLKLDKINLTDNYVL	60
1	MSLVNRKQLEKXANVRFRVQDEYVALDLAEYHNHSENTVVEKYLKLDKINLSTDYVI	60
61	NTYKSGRNKALKKFKYELTWELVLELKNSLTPVEKNLHFIMWIGQINDTAINYINQWKD	120
61	DTYKSGRNKALKKFKYELVIELELKNSLTPVEKNLHFIMWIGQINDTAINYINQWKD	120
121	VNSDYTVKVFYDSNAFLINTLUKKTIVBSATNNTLSPRENLDPEFDYKNPKYKRMELIY	180
121	VNSDYVNVVYDSNAFLINTLUKKTIIESASNDTLSPRENLDPEFNHTAFPRKQMIIY	180
181	DQKHPIIDYKSOEENPEPIIDNIIKTYLSNEYSKDLEALNKYIEESLNKTIANNNDI	240
181	DQKQFNINYYKQAKBENPDLDIDDIKTYLSNEYSKDIDELNAYIEESLNKVTENGNDV	240
241	RNLEKFADEDLVRLYNOELVERWNLAASDILRISMLKBDGVYLDVILPGIOPDLFKS	300
241	RNFPEKTVGFVNIYEQESVERWNLAGSILRLVAILKNIGGVYLDVMDLPGIHDPDLFXD	300
301	INKPDSI-TNITSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVORSFESALASKSDKSI	359
301	INKPDSVKTADVDEEMQLEAKMKHKEYIPEYTSKHPDLDVEEVQSFESVASKSDKSI	360
360	FLPLDDIKVSPLEVKIAPANNSVINOALISLKDSYCSDLVINOIKNRYKILNDNLNPSIN	419
361	FLPLGDIENVSPLEVKIAPAGSIIINOALISAKOYSICDLIKQIQNRYKILNDTLGPIIS	420
420	EGTDENTMTMKI FSKOKLASISNEDNMFMWIKITNLYKVGFPADVRSITINLSGPGVYTGAYQ	479
421	QGNDENTMNNFGSLSGIANEENISFIKIGSVLRVGFPYEANTWITLSGPTIYAGAYK	480
480	DLLMPKONSTNIHLLPELRFNPEPKTKISQLTEQEITSLWSFNQAPAKSQFEYKKGVF	539
481	DLLTFKEMSIDTSLSSBLRNFPEPKVNI SQATEQEKNSLWQFNBEERAKTIFEEYKKNVF	540
540	EGALGEDNDLDPAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVLOQDKLISYEASCNLPS	599
541	EGALGEDNDLDFSQNTYDKLEYLLEKISSTKSGSEGGYVHYIVLOQDKLIEYAACNLFA	600

Qy	600	KDPYSILYQKNTGSETATAYYYVADAIEKIDKYRIPQISNKRNIKLTFIGHGKSEFN	659
Db	601	KNFYDILFORNIETSEVAYYNPTDSEIQEIDKYRIPDRISDPKIKLTFIIGHKAEFN	660
Qy	660	TDTFANLDVDSLSEIETILNLAKADISPKYIEINLLGCNMFYSYVAEETYPGKLLKI	719
Db	661	TDIFAGLDVDSLSEIETATGLAKEIDISPKSIEINLLGCNMFYSYNVEETYPGKLLRV	720
Qy	720	KDQVSELMSPISQDSITVTSANQYEVIRINEGKREILDHSGKWINKGESIIKDISKEYIIS	779
Db	721	KDKVSELMSPMSQDSIIIVTSANQYEVIRINSEGRELLDHSGEWINKGESIIKDISKEYIIS	780
Qy	780	FNPKENKIIVKSKVLHELSPLLQEIIRNNANSSDIDLEKKVMLTECRINVASNIDRQIVEG	839
Db	781	FNPKENKIIVKSKNLPFLSPLLQEIIRNNSSSDIELEKVMLAECENIVSNISNIEETQVVE	840
Qy	840	RIEAKNLTSDSINYIKNEPKLIESDSILYDLKHQNGLDSDSHFISFEDISKTENGFRIR	899
Db	841	RIEAKSLTSDSINYIKNEPKLIESISeALCDLKQKQNELEDSHFISFEDISDETFGSIR	900
Qy	900	FINKETGNSPIETKEIIESEYATHISKEISNLIKDTIFDNVNGKLVKKVNLDAAEVNTL	959
Db	901	FINKETGESIFVETEKTIIFSEYANHAITEEISKIKGTIFDTVNGKLVKKVNLDTTHEVNTL	960
Qy	960	NSAFFQSLLEYNTTKTESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETI	1019
Db	961	NAAFFQSLLEYNSKESLNSLVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETI	1020
Qy	1020	D 1020	
Db	1021	D 1021	

RESULT 4

A37052 toxin A - Clostridium difficile  
 C:Species: Clostridium difficile  
 C:Date: 31-Jan-1992 #sequence revision 31-Jan-1992 #text\_change 24-Sep-1999  
 C:Accession: A37052; A60991; S21897; S2437; S08638  
 R:Dove, C.H.; Wang, S.Z.; Price, S.B.; Phelps, C.J.; Lyster, D.M.; Wilkins, T.D.  
 Infect. Immun. 58, 480-488, 1990  
 A:Title: Molecular characterization of the Clostridium difficile toxin A gene.  
 A:Reference number: A37052; PMID:90129305; PMID:2105276  
 A:Accession: A37052  
 A:Molecule type: DNA  
 A:Residues: 1-2710 <DOW>  
 A:Cross-references: GB:M30307; NID:g144925; PIDN:AAA23283.1; PID:g144926  
 R:Wren, B.W.; Clayton, C.L.; Rabagchali, S.  
 FEMS Microbiol. Lett. 70, 1-6, 1990  
 A:Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and  
 A:Reference number: A60991  
 A:Accession: A60991  
 A:Molecule type: DNA  
 A:Residues: '1', 1894-1899, 'K', 1901-1910, 'Y', 1912-1919; 2054-2074; 2096-2116; 2138-  
 A:Cross-references: GB:X17194  
 R:Eichel-Streiber, C.; Laurenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sau  
 submitted to the EMBL data Library, July 1991  
 A:Description: Comparative analysis of Clostridium difficile toxins A and B.  
 A:Reference number: S21894

A;Molecule type: DNA  
A;Residues: 1-154 <EIC>  
A;Cross-references: EMBL:X60984  
R;von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Mol. Gen. Genet. 233, 260-268, 1992  
A;Title: Comparative sequence analysis of the Clostridium difficile toxins A and B  
A;Reference number: S22434; MUID:9223124; PMID:1603068  
A;Accession: S22437  
A;Molecule type: DNA  
A;Residues: 1-92 <VON>  
A;Cross-references: EMBL:X60984  
A;Note: The four fragments shown in reference A60991 correspond to four types of repeats ordered ABCCCDABCCDABCCDABCCDABCCDABCDABC



Qy	872	LKHQGLDDSHFISFEDISKTENGFRIPINKETGNSIFITEKEKIFSEYATHISKEISN	931
Db	878	FYQKNISNNMIILFDSIIKSYNNVWKLANKITGETSVIKTYSLSLWFTNKYKKIVDD	937
Qy	932	IKDITFONVNGKLVKQVNLDDAAHEVNTLSNAFFQISLIJEYNTTKESLSNLSVAMKVQVYA	991
Db	938	IKGIIVKIDINGEFIKADFEIEQNPSSLNSAMLQMLLIDYKPYTILTNMTSLKQVYA	997
Qy	992	QLPFTGLNTITDASKVVELVSTALD	1016
Db	998	QIFQISGAIQEATEIVTIISDALN	1022
RESULT 6			
T00296			
toxin B - Escherichia coli plasmid p0157			
C:Species: Escherichia coli			
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 17-Nov-2000			
C:Accession: T00296; T42195			
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H. C.			
S.; Shinagawa, H.			
DNA Res. 5, 1-9, 1998			
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an ente			
A:Reference number: Z14127; MUID:98290540; PMID:9628576			
A:Accession: T00296			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-3169 <NA>			
A:Cross-references: EMBL:AB011549; NID:G4589740; PIDN:BA031815.1; PID:G3337056			
A:Experimental source: strain EHEC O157:H7, substrain RMD 0509952			
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, P.R. J.			
Nucleic Acids Res. 26, 4196-4204, 1998			
A:Title: The complete DNA sequence and analysis of the large virulence plasmid			
A:Reference number: Z22068; MUID:98391744; PMID:9722640			
A:Accession: T42195			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-236, 'D', 238-1887, 'I', 1889-3169 <BUT>			
A:Cross-references: EMBL:AF074613; PIDN:AAC70163.1			
A:Experimental source: strain EDL933; serotype O157:H7			
C:Genetics:			
A:Gene: toxB			
A:Genome: plasmid p0157			
A:Note: L7095			
C:Superfamily: Escherichia coli probable cytotoxin			
C:Keywords: cytotoxin			
Query Match 7.9%; Score 408.5; DB 2; Length 3169;			
Best Local Similarity 20.7%; Pred. No. 1.2e-07;			
Matches 255; Conservative 200; Mismatches 449; Indels 325; Gaps 55			
Qy	14	YVKPRI---QDEYVAILNAL---EYHNMSSSVVEKYKLKDINLNTDNYLNTYKKSGR	68
Db	160	YIKIRKTRGASDQTITQSLIINELLNGVDNRNTI-PPQKISELNDIHSYENWMIKNSR	218
Qy	69	-----NKAALKFKE-----YLTWELVELKNSLTPVEKNLHFTWIGS	105
Db	219	KGIELVKQGLLSLNVNNGKSLNSASKIINLLGIEYOSHKV-DIEPFIAVWVAG	277
Qy	106	QINDTAINYNQWQDVNSDYTVKFVDSNAF-----LINTLKKT-----	144
Db	278	APPDNTSYITAFNLTYKYDYTYLLWIDPNAFGAAKFSGILKNIAMYIMRLRRTNPHLA	337
Qy	145	-----IVGSAVNTTLE--SFRENLDNPFDY-----NKFYRKMEIYYDKQKHP	186
Db	338	EMNEVILKIQNIQNETIEPKETRELRKLENNRYKSLTSETKEKFNVPFLESIMIGMDNY	397
Qy	187	IDYYKQSQEENPERT-----IDNIITK--TYLSNEYSKOLEALNKYIEESLNTKITTANNG-	238
Db	398	FTYCISNGISNTDDLSRLDFTLNVLUKLSPEVQNDPKASTVEKNKRDIDLKNTISQKFGDR	457
Qy	239	----DIRNLEKFADEDLVRLYNQELIVERWNLAASADILRISMLKEDGGVYLDVILPGIQ	294

Db 458 FQLRDINTLESFKKQDQFFYQOQMLRNWYAAASQVRINILKEVGGIYTDILPAYS 517  
Qy 295 PDLFKSINKPDSITNTSWEKILBAIMKYK--EYIPG--YTSKNFMDLDEEVORFESAL 350  
Db 518 DKVSQILNE-KSDDKRFPPDLRLRISISLSIKGKYSIKH-DGLDETTLQNLNLT 575  
Qy 351 SSKSDKSEIFPLDDIKVPLEKIAFANNSVI-----NQALISIKOSY 394  
Db 576 -SEIEK--LIIDDY-FKPVETKVVRDFTKIFKYQKWNTWIRGNWNNFMTLTKGSK 629  
Qy 395 CSDLVINOIKNRYKILNDNLNPSINEGTDFTNTMKIFSDKLASISNE-----DN 443  
Db 630 CIDFILSGQKQVLEL-QRIRDNISYNLFYTT-----EDLAKLNNAVGIGIPAKKYLEH 683  
Qy 444 MFMKIKITNLYKVGFPAPVRSINLSGP----- 471  
Db 684 GLF-----SEYRQDGTIPYVVTNLSGPDMMIMROMKKYKSLGRIGEVHIKONKLSVNF 739  
Qy 472 -GVYTGAYODLWF---KONSTNIHLLEPE-----LRNFEPKTKISOLT---EQEIT 517  
Db 740 LGVYASSNKONKSPWLNPNVSVGINDITPDDESSWAVRNNDINKILFEKINCHVPEKLP 799  
Qy 518 SLWSFNQARAKSQFEYKGYFEGALGEDNDLDFAGNTVLDKDYVSKILSSMKTRNKEY 577  
Db 800 SLV-----YEIDRSRFFQGW---DNKSIRKHVTEINKDLI--KQINLLTSSNID 843  
Qy 578 IHVIVLOQDKISYEASCNLFPSKDPYSSILYKQNIESTAYYYVADAEIKDYRIP 637  
Db 844 VKLLIKL--DRELYAISSKI-----DNPLARSIRTQLQANVYVTSNTPENTINFI 895  
Qy 638 YQISNKRN-----IKLTPIHGKSEFNTDTANLDVDS-----LSSEIE 676  
Db 896 YDFYRKQDDLSAIIK-----FSRNDADTKIIWNVSMEKVFREVLS 941  
Qy 677 TIINLAKADISPKYIEINLLGCMFYSIYABETYPGKLLKIKORVSELMPSISQDST- 735  
Db 942 CVLRSKKVD--SYINEN-----KKNLSKEDA--GALRDYAKLKMKEFLSMDDGKY 989  
Qy 736 -TVSANQYVRINEEKRILDHSGKWINKESIIKDISKEYISFNP----- 782  
Db 990 KIITTNAY---IKERDKL-----SGIYIENSIISGHESFDIIRSNQHEWGLSTVEOF 1041  
Qy 783 KENKIIVKSKYLHELSTLLQEIARNANSIDILEKKVMTLCEINVAS-----NIDRQIV 837  
Db 1042 KKEFEVVKSE-ISSAKSIFDDIKK-YITDPETKRNVLVHOLDSDIKERIAFLDISHYAY 1099  
Qy 838 EGRIEAKNLTSDSINYIKNEPKLIESI-----SDSLYDLKHQNG 877  
Db 1100 PGSLLLEKLQLS---GVVFSIDINIIAEYLLASVGVSHVSHGVVYVAPSDKLELLRRHT 1155  
Qy 878 LDDSHFISFEDISKTENGFRIRINKETGNSI---FIETEKIFSEYATHISKEISNID 934  
Db 1156 KGNSEWI--EKITP-----YVYDILSDNVGNVLRPLSEBQKILNDIKLEISKSVS--E 1206  
Qy 935 TIFDNNVGLKVKVNLDAAEVNTLNSAFFIOSLIEYNTTK-----ESLSNLSV--- 983  
Db 1207 QYPMKLTQKSSVIGIKYVDFDRYNNELFLSLPINQNLTLPFMYRYFEMLIDHIGITE 1266  
Qy 984 --AMKVQVYAQLFSTGLNTITDASKVVEL 1010  
Db 1267 NKANREPIYKPSLSLNDFLINDERVNL 1295

RESULT 7  
DB1702  
adherence factor TC0439 [imported] - Chlamydia muridarum (strain Nigg)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: DB1702  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: DB1702  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3225 <TET>  
A:Cross-references: GB:AE002312; GB:AE002160; NID:g7190482; PID:AAF39293.1; PID:g719048  
A:Experimental source: strain Nigg (MoPn)  
C:GeneticS:  
A:Gene: TC0439

Query Match 7.0%; Score 362; DB 2; Length 3225;  
Best Local Similarity 19.8%; Pred. No. 6.7e-06;  
Matches 254; Conservative 212; Mismatches 414; Indels 404; Gaps 63;

Qy 2 NLVKAQLOKMYVVKFRIQEDSEYVAIL-NALIEEYHNS-----SSVVEKYLKLDINN 54  
Db 196 NILEKLSLTQ---EQIKHDNLISNLKEANRYSDLNKRNKRKQSLLRQAKILD--E 249  
Qy 55 LTDNYLNTYKSGRNKALKKFEYLTMBVELKNNSLTPVEKNLHFVIGQINDTAINY 114  
Db 250 ILSQTKSTEERASNSVMTTIKEFTSHRV-----PVEKNHIGIWIAGSPPGTDEY 300  
Qy 115 INQMKDVSDDYVVFVYDSNAP-----LINTLKKTIVESATNNLTSPRENLDNPFVYNK 170  
Db 301 IKLFLHTYPEFSFLFWDKTAYGAAKFSSTLKRIAFDAVN---SLREATPEPVKQFVQ 356  
Qy 171 FYRKMEIILYKQKHFIDYYSQIBENPEFIIDNIKIYLSNEYSKDL----- 218  
Db 357 RYDK-LKSYDTSRDFDE--KQRLSEQLVELYDNY-----NKFSKEIQSNFVLLLHEM 407  
Qy 219 -----EALNKY-----IEESLANKI 232  
Db 408 ITIQDSFFNYCOLKGVGAITDETRIEYLEKVLKVEEDLSHYKETIKNKSESIEKLVKEI 467  
Qy 233 TANNG-----NDRNLEKFADEDLVRLYN--QELVERWNLAASDILRISMLKEDGGVY 284  
Db 468 NDSTGRERWIKIDRLAKSL--QDLTNSYNVETEMLLRWNYAAATDQLRMVLMKEYGGIY 525  
Qy 285 LDVDILPGIQDPLFKSIN-----KPDSTNT--SWEMIKL-----EAIMKYKEIPLYGT 331  
Db 526 TDLTIMPOYSQDVLQKINDVGSGSRFFEHDKLRRTLTSLFAALKLGSGKQTTVSFEAKKAMT 585  
Qy 332 SKNFMDLDE-EVORSFESALSKSKSIFLPLDDIKV---SPLVKI---AFANNSVI 383  
Db 586 LPTFFLQKQSISEIFKYLETQAKSLFQPMQVTVVVDPMPIQORVHKWQTHNVRGL 645  
Qy 384 NOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINEGTDFTNTMKIFSDKLASISNEDN 443  
Db 646 NGLMMAHKDSAVVDARIAQRAAYDEM--ALRQNVVSGEFFRSL-----GDLEHVRNKN 699  
Qy 444 M-MFMKIKITNLY-----KVGFPAPVRSINLSGPGVYTGAYQDILLM--FKDNSTNI 491  
Db 700 IGGYLA--NYLGGSLFFDFRQDSVIPGAISTLGISGP-----DIIMDTMSDYFTNL 749  
Qy 492 -----HLEPELRN-----PEPPTKISOLT-----EOEIT---SLWSFNQA 525  
Db 750 GPVGEDFLYEGKLGKAAFLGAYQAKTPKGELTYDLHLPLSIGANDVTPADASTWCETRQ 809  
Qy 526 RAKSQF-----EYKKGYFEGALGEDD-----NLDPFAQ-----NTVLDK 559  
Db 810 HCAAEALLLSDSISDSDEHPKGRRRVRNPNDFSKLWKEAQGLSDDFADLLPRFNLLIES 869  
Qy 560 DYVSKKILSSMKTRNKEYIHYIVQLODKISYEASCNLFPSKDPYSSILYKQNIESTAY 619  
Db 870 SALDIHTLSAL---DRDIQHLFTKVQ-----KDPVASVA----- 900  
Qy 620 YVYVADAIEKIDKYR-IpyOISN-----KKNIKLTFIHGKSBF----- 658  
Db 901 ----VFSLOQLAEMRAIPFPPIRNQVHLPEQAQHFADWKAKAIQIYLHSHFQTEVWIY 957  
Qy 659 ---NTDTFANLDVDSLSSEIETILNKAADISPKYIEINLLGCMFYSIYABETYPGKL 715

Db 958 SSTHTQIVFGKOLLAAVERAAAKSL-MSDHPDPLIT-----SYLKRYKTQSHLGLV 1007  
Qy 716 LLKIKDRVSELMPSISODSITVSANQYEVINEE---GKREILDHS-GKW--INKEE--- 766  
Db 1008 TFDQEDFELMVDIAEP---ELHKQLDKIEQNSGLYSHVEHSLGELWKLKSEERKS 1064  
Qy 767 ---SIIOISSKEYISFNPKENKIIVKSKY-----LHELSTLLQEIIRNNAN 809  
Db 1065 KPLKILKETFEEREDSQOQKHTFEELYEKRHQVRKDPKAKIQLITVQB-SQRVQ 1123  
Qy 810 SSDID-----LEKVMLTECEINVASNIDRQIVEGRIEBAKNLTSINIKNEPKLI 862  
Db 1124 AQDIDITFAHKPFYQDLMKDQYAFEDISVITKYL-----LASDGVSGIITDPF 1173  
Qy 863 EGISDSLYD-LKHQGLDSDHSPISFEDISKTEGFRIRF--INKETGNSFIETEK--- 915  
Db 1174 PPPSKQLIDAMKQSLG-----EDFGLHYTLQWYDMLSKET-NSVTSEQAQKQLP 1223  
Qy 916 ----EIFSEYATH-----ISKEISNIKOTIFDNV----- 940  
Db 1224 QKLHEKLEGYTHDLLIPIDGVSALGLRPFSTEEGKVSDRVLTSLAPGVPSASYAMTS 1283  
Qy 941 -----NGKLKVKVNLDAHEVNTLNSAFFTOSLIEYNTTKESLSNLSVAM-- 985  
Db 1284 YLYGLFLITKQIQSGRLTHEI---LKERLQTYGGAYFI-----NESKIDVILLALS 1331  
Qy 986 -KVQVY---AQLFSTGLNTITDAS 1005  
Db 1332 KKAQISLIDAHKALTGFSFSSEAS 1355

RESULT 8  
T28676  
rhoptry protein - Plasmodium yoelii (fragment)  
C:Species: Plasmodium yoelii  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T28676; A45521  
R:Singha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass  
A:Reference number: Z20507; MUID:97077455; PMID:8920022  
A:Accession: T28676  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2401 <SIN>  
A:Cross-references: EMBL:U36927; NID:G1041784; PID:G1041785; PIDN:AAB41263.1  
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KE>  
A:Cross-references: GB:M34281

Query Match 6.7%; Score 350; DB 2; Length 2401;  
Best Local Similarity 20.0%; Pred. No. 1.3e-05;  
Matches 261; Conservative 202; Mismatches 391; Indels 454; Gaps 63;  
Qy 27 ILNLEEVHNMSESSVWEKY-----LKLKDINNLTNLYTYKSGRNKALKKPK 76  
Db 62 LINKVDSGNIIEENDIDNFDLSKPKNNFKLEGEINGVFENKNDYK---NNKVLDPK 118  
Qy 77 EYLTVMEVLKGNLSTPVEKNLHFIWIGGQINDTAINYINQWKNVSDYT-VKFFVDSNA 135  
Db 119 DTMKIILLI-----IQYNPEKGLDAMTKLKNBSIQK 153  
Qy 136 FLINT-LKTYIVESATNTLESFRENLDNPEFDYKFRKREIYD----- 181  
Db 154 FVINQIKQPKDKSTYDEKKEGFESL---ELAKN-WKKKKLEIITELKKNQNEETVOLDI 209  
Qy 182 KQKHFDYKQSIENPEPIIDNIITKYLNS-----EY-----SKDLEALNKYI 225

Db 210 KIRELIKQKDIIEBQK---IVNDKLKELNKKIKETIEYIKKAVDLKKEIEKDNVYI 266  
Qy 226 EESLN---KIT--ANNENDIRN-----LEKPAEDLVRLYNQ--ELVERWNLAASDIL 272  
Db 267 DELAKBPPYQITKYEKKEINEYNTIKSDFDKIYVGDIEQLYNEMFSVVOESNIEHNKT 326  
Qy 273 RISMLK-BDGGVYLDV-----DILPGIQPOLFKSINK-- 303  
Db 327 ELTLTKIDNVYNNIQNMETETVKSHLKNIETNNKLSETILDIILKIYGEITNELNKT 386  
Qy 304 -----PDSINTSWEIMKLEA---IMKYKEYIPGY-----TSKNFMDLD 339  
Db 387 EDFKNKEKGLSKNIDEYAKENVQLNVYKSNILEIKKHYNDQINIDNIKEKAQYDQPK 446  
Qy 340 EEVQSFESALSSKSKSEIFLPDDI----- 366  
Db 447 EHMKTIPPEWYKQPSIEIKIMKDEFUSKVNKNYDFKVKYKVESEHNKTELTNLIK 506  
Qy 367 -KVSPLVKIAPAN-----NSVINOALISLKDSY-----C--SDLVINQI 403  
Db 507 TEVSDEEIK-KYENKFNDSKSLINETKKSIEEYQININTLKKVDDYIKVCLNTNELINC 565  
Qy 404 KNRKILNDNLPNSINEGTDFTNTWKIPFSDKLASI----- 438  
Db 566 HNKQTTLKDLQNLQIKTIKETNSDKIYTKFENILTDKTELETKFTGLSLNNHESNNK 625  
Qy 439 -----SNEDNMFM-----IKITNLYLKVGPAPDVRSTINUSGPOV 473  
Db 626 ELLTYFYDLKANLGNKNKNMLYKQPNKEKAVEDIKQNV-----DINKIVSNIETI 678  
Qy 474 YTGAVODLLMPKDNSTNIHLEPELRNFPFKTKISOLTEQOEITSLWSFNQARAKSOPEE 533  
Db 679 YTSIYN---INEDTENEIG-----KSIELNTKVKLEKAVNTNL-----NEIKE 720  
Qy 534 YKGYFEGALGEDNDLPAQNTVL--DKDYVSKILSSMKT-----RNKEYIHYIQLQ 586  
Db 721 KLDYDFQDFGKEKNIKYPDENKIKQNDITLQKIDKSITELTEIKKNS--NHIDEIKG 778  
Qy 587 --DKISYEASCNLSKDPYSILQKNEG-----SETAYYYVADAEIKEDIKYRIP 637  
Db 779 QIDKLKKVPNKTMFNEPKK---TEKKIENIVEKIDKKKNY-----KEIDK--LL 824  
Qy 638 VOISNKRNIKLPF-----IGHGKSEFNTDTFANL----- 666  
Db 825 NEISKIENDKTSLEKKNINISYGS-----LGNLFQOIDEKKKAETIKAMEAYID 878  
Qy 667 DVDSL---SSIEFTILNL---AKADISPKYIEINLLGCNMFYSY----- 706  
Db 879 DLNLIKKSQIEKEMNINMDIKMDI---HKEMKALNISHDDYKIYHTTSKNHEEKISDI 935  
Qy 707 -----ABETYPGKLLKIKDRVSELMPSISODSITVSANQYEVINEEGRREIL 755  
Db 936 RNSLKIIQDFSEESYINDIKKELEKNVLE-----SQNNNT-DINQYLSKI--ENIYNIL 987  
Qy 756 DHSGKWINKESIIKDISKEYISFNPKENKIIVKSKYLHELSTLLQEIIRNNANSIDL 815  
Db 988 K-----LNKIKKIDKV--KEYTDEIEKNNK-----KINAEUS-----NSEKIIQL 1027  
Qy 816 EKKVMLTECEINVASNIDRQIVEGRIEBAKNLTSIN-----YIK----- 856  
Db 1028 KENSLKESQSKIKSTIDNNVSECINKITNLKTYIVNEKNKNINYFKNAEYQNVSIN 1087  
Qy 857 -NEFKLIESISDSLYDLKHQGLD--DSHFISFEDISKTENGFRIRFPINKETG-NSIFIE 912  
Db 1088 FNNIEMADTKSQYILNINKNNGTNTDYNIKELKEHKKKSNVYK-----DEAGKNTQEI 1142  
Qy 913 TEKRIEFSVATHI-----SKEISNIKOTIFDNVNGKLVKKNLDAHEVNT----- 958  
Db 1143 KKNKELFKEYOEVTVLLNKYYAVELDNKPKDT-KNYSQIIEKIK--DAHNTFTSOADKS 1199  
Qy 959 ---LNSAFFIQLSLEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITD 1003



Db 1200 EKQWEIKNEQRIEDEVAKONNSKAI-LDIQLSVEFFKIFLKIKD 1246

RESULT 9

T28317

ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus

C/Species: Melanoplus sanguinipes entomopoxvirus

C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C/Accession: T28317

C/R/Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.

J. Virol. 73, 533-552, 1999

A/Title: The genome of Melanoplus sanguinipes entomopoxvirus.

A/Reference number: Z20484; MUID:99102612; PMID:9847359

A/Accession: T28317

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1127 <AFO>

A/Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717

C/Genetics:

A/Note: MSV156

Query Match	6.7%;	Score 345.5;	DB 2;	Length 1127;
Best Local Similarity	21.9%;	Pred. No. 7.3e-06;		
Matches	248;	Conservative 167;	Mismatches 368;	Indels 349; Gaps 54
Qy	9	LQKVVVVKFRIQDEQVAILNALAEYHNHSSSVVEKYKLKQIANNLTNDYLNATYKKS--	66	
Db	61	VEKIFVHMQKINDNIILQVILBYNNEIKCKENKFPCK--NPL---YNTYKKYU 115		
Qy	67	--GRNKALKFKFEYLTMEVLELKNNSLTPVEKNLHFHWIGGQINDTAINTINQKDVNSD 124		
Db	116	IYDLDVVEEKOKELVIN--IEQKN---AVDK-----INDIKVNNIHSNDETI 159		
Qy	125	YTVKFVYDSNAFLINTLKTIVESTATNTLGSFRENLDNPEFDVKNFYKRMELIYDKOK 184		
Db	160	ITGK---ETLIDILANKLK-LVSSDEKQLIEQIKYKNNKIEIEPKN----- 201		
Qy	185	HFIDYYKSOIENPEFIDNIITKTVLSNEYSKOLEALNKYIEESLNKITANNNDIRNLE 244		
Db	202	-----IDNVOK-----EINKQDELNKLDDSKKEFIKKQ-----EELN 235		
Qy	245	KFADLVLRYNQELVERWNLAASDILRLSMLKEDGGVLDVDVILFGIOPDLFKSINK 304		
Db	236	KTIKK---QBELIKKLN-----DKNEINFND-----EQKLLDQI 268		
Qy	305	DSITNTSWEMIKLEAIMK-YKEYIPGYTSKNFMDLDEBVSQSPESALSSKSDKSEIPLPL 363		
Db	269	NSKINTLNENIK--GYMNLYTE-----TKNKISNLQNEILNK-DSTIKSLDERQKLLDEL 320		
Qy	364	DLIKVSPLEVKIAPANNVIN-QALI--SLKDSYCSDLVINOIKNRVKI-----LNDN 413		
Db	321	DK--NINNITSLYKNGKTIKITNQLESLSLTFNNANININELSKSKTKLPNDIOKLND 379		
Qy	414	LNPSINEGTD-FNTWKIPSPDKLAS-----ISN-----EDNMFMKITNLYKVGA 459		
Db	380	ITEQNNKIITDFNNSTRIFPEKCLDTEYKKIDDIKNNLQKLESYKKIDQETYYKNKN 439		
Qy	460	PQVRSITNLSPGV-----YTGAYQDMLPFKDNSTNIHLSPER 499		
Db	440	KEYNDIIEELKNNLQKLEENKKIDQETYYKNKINKEYNDIIEELKNN--NLQKLEENK 497		
Qy	500	NPEFPPTKLSIQTEQETISLWSFNQ-----ARAKSQFEEKYKGYEGALGED 546		
Db	498	NINDKUTLKNDIEN--TEL--FNKLNTSDFPKRSREIAKLNTYEQRLKDLLENIKNTN 554		
Qy	547	DNLDPAQNTV--LDKDYVSK-----KILLSMKTANKEYIHYIVQLQGDQKSYEASCN 596		
Db	555	ELMKLSDNKLSSLEQLYDSKKNILDGIDKIYNSLKEKN-----DKI----- 595		
Qy	597	LFSKDPYSILYQKNIEGSETAYYYYVADAEIKIDKYIPYQISNKNRKITLFTIGHGKS 656		
Db	596	---DBYFS-----NIEKFDIYVNIENK-----FIGNLDS 621		

Qy	657	----	EFNTDTF----	ANLDVDSLSSEISFI	TLNLAKADISPKYIE	INLLGNMPSYSIYAE	709	
Db	622	I	INKIINNDQFKEYINS	KIDSKSNELSTMFD	----	DFNAKNQIASITNNIENIS	672	
Qy	710	TYPGKL	LLKIKIDRVSELMPS	ISQDSITVSANOYEVR	INEEGREILDHSGKW	INEESTI	769	
Db	673	----	NKIKD-----	LNEFIIS-NEDSSKE	LDEIRKY-KQOFDKI	706		
Qy	770	KDISKEYISFNPKEN	KII--VKS	KYLHELSTLLQEI	RNNANSSDIDLEKKV	MLTECEI-	826	
Db	707	KDAMNTEVKSFENTL	QKDISKSN-INEL	NAYDIINTKAN----	DLDDKLNNYGSSEPK	761		
Qy	827	----	NVASNIDRQIVEGRI	EERAKNLT-----	SDSINYIKNEF--KLIE	-----SISDS	868	
Db	762	NLYNNASDLDTIO	KNNDEKVKQLNLEY	EKKKNQNSIEINDI	VNNEFKELIKENNT	ETNKS	821	
Qy	869	LYDLKHQNGLDSD	HFISFEDISKTE	NGPRIRFINKETGN	----	SIFETEKEI	FS----	919
Db	822	LNELLTDDNDIND	KIFPKYKELNK	ISTNNLLKIYKNEID	NVNEKUSIVLENQ	OFINSFUSI	881	
Qy	920	----	EVATHIS----	-----	KEISNIKDTFDNVN	GKLVKKV	948	
Db	882	EFNQSITSHINFL	NLTLAGINDVLN	KLNLKIMADTT	TRGDTNIRDEIK	QNSISENIKQ	941	
Qy	949	NLDAAEHVNTLNS	AFQISLIEYNTT	-----	KESLNLSVAMKVQVY	990		
Db	942	KFNEKNEKD	KLKLSFNDKLN	KYNISAGYTEYNNIE	HECLKYLYIAVS	DQEY	993	
RESULT 10								
Tl8427								
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)								
C/Species: Plasmodium falciparum								
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000								
C/Accession: Tl8427								
R/Lawson, D.; Bowman, S.; Barrell, B.								
submitted to the EMBL Data Library, August 1997								
A/Reference number: Z18935								
A/Accession: Tl8427								
A/Status: preliminary; translated from GB/EMBL/DBDJ								
A/Molecule type: DNA								
A/Residues: 1-3724 <LAW>								
A/Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB1104.1								
C/Genetics:								
A/Introns: 307/1; 1545/2								
A/Note: C0335c								

Query Match	6.4%;	Score 330;	DB 2;	Length 3724;
Best Local Similarity	21.8%;	Pred. NO. 0.00013;		
Matches 228;	Conservative 176;	Mismatches 351;	Indels 292;	Gaps 58
QY	20	QEDYVAIILNALBEEYHNMSBSSVVEKYLUKLIDNLTNYLNTYKSGRNLKALKKPKFYL 79		
Db	618	EDDEVYHVLK-----ENLKEDA-----NEYNDKEN-----KNNKTBEIKSKNYL 658		
QY	80	-----TWEVLEL--KNNSLTPVEKULHFIWGGINDTAIYNQWKDVSNDYTKFVYD 132		
Db	659	ENEKRTLEELKLRGNIFKDEKY-----NSLGEVINEI-----QINEENKINDIQD 707		
QY	133	SNAPLINTLKTIVESA--TNTLTLESFRENLDNDFYKFKRMEIYYDKQKHFIIDY 190		
Db	708	G-----NISKQIIQSSRTNDTFNIDISLD---DLEKEKRK-----KSOHFIDNL 753		
QY	191	---KSOIEENPEFIDMI IKTY---LSNYSVKDL-----EALNKYTEESLUNKITANNQ 237		
Db	754	VKADNVEISENINKIDNNINNIYDESINNIYDESINNIYDESINNIYDENI 813		
QY	238	NDI--RNLEKFADEDLVRLYNQ-----ELVERWNLAASDIILRISMLKEDGGVYLDV 287		
Db	814	NNIYDENINNIYDENINNIYDEGINKICDDNILENKNIKTINDIYQV---EENNESIE- 868		
QY	288	DILPGIQDPLFKSINKPDSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVORSE 347		



Db	869	-----KNBLMISLAKD-----INNTYMPK-----ENVDIFINKIKR--E 901	Matches 255; Conservative 218; Mismatches 388; Indels 407; Gaps 65;
Qy	348	SALSSKSKSEIFLPLDDIDKSPLEVKIAPANNVSINOALISLSDSYSCSDLVINOIKRY 407	Qy 3 LVNKAQLQKVVYKFRIOEDEVAILNALALEEYHNMSSESVVEKYL--KLKDIINNLTNDYL 60
Db	902	SIL--KIDK-----NIK-----DNN---NDDEYMDNPYENDFIIN--HKM 935	Db 8 LVNLFVIVIFV-YR-QYDKRSRALDKIKFVDITKVN--LEDIFIEDTKTEINDLAVD-M 63
Qy	408	KILNDLNP-SINEGTFNTWKIFSDKLAS---ISNEDNMF-MIKITNYLKVGFAPDV 462	Qy 61 NTKYKSGRNKALKKPEKYLTMVELELNKNSLTPVEKNLHFIIWIGQIINDTAINYNQWKD 120
Db	936	EITNKELDPLEINTQNEFIENLDIKKKYVNDHFNDADKMFYEMNKILN-----KDM 988	Db 64 BAYQRSS-IEIUKIEE--VQOKIKVNSDPAVEKKIAY-----HDSMLKOLDDEMTF 113
Qy	463	RTINLSGCVVTGAYQDILLMFKDNSTNHLPELRFEPFKTKISQITEQIETSLWSF 522	Qy 121 VNSDYTVFVDSNAFLINTLTKTI-----VESATNNTLESF-----RENLDPEF- 166
Db	989	KKN-----KEQFFKTDFTGSLQSHKIKKYNKGEKHDKNKEEKILYDE 1035	Db 114 KVQENIQLOVDGK--IVDKLSKTLKGFNTQIDSVESNLNSVLEKDFKANKENLESIKTA 171
Qy	523	NOARAKSQPEYKKGVEGALGED-DNLOFAQNTVLDKDYV-----SKKIL-----SSM 570	Qy 167 DYNKYRKRMEIY--DKQKHFDYKSOIENPEFIIDNIK-----T 208
Db	1036	NOV-----YSVLSDHKIEQDQDIHSIQTNICDENNIQIENESKKGVRISGTM 1087	Db 172 SWEKEDTNIKELVFKIDNLNKEISYKOLANIEBERKNDILVKNEKLDLDFSDLEKVE 231
Qy	571	KTRNKKEYIHYIYVLOQD---KISYEASCNLFKDPYSSILYQKNIEGSETAYYYVADAE 627	Qy 209 YLSNEYSKDLREAL-----NKY--IEESLNKITANNNDIRNLEKFADEDLVLRYNOELVE 261
Db	1088	ENKNDMENKNDMEKNDMEKNDMEKND-----MEKKN-----DME 1130	Db 232 FNIQKYSKEIESSFNFYENKYKLIENSIELIMESVKNKINEKEDF-----ILNRLNEELQN 287
Qy	628	IK-EDIKYRIPQISKNRIKLTIFGHGKSEFTDTFANLVDVDSLSSEIETILN-----680	Qy 262 RWNLAASDIL-----RISMLK---EDGCVYLDVDI---LPGIQDPLFKSINKPISITWT 310
Db	1131	KKNDMEKKN---DMENENME-----NKSDIENENOMNEYKNNMKENKMIENNTIENN 1181	Db 288 KP-----KDIIVYDDRSKGIKDKLEDKLVLDNEISSMSSSFKNVYSRINSLB-----337
Qy	681	LAKADISPKYIEINLGCNMFYSYIABETYPGKLL---LKIQRVSLEMPISQDSIT 736	Qy 311 SWEMIKLEAIMKYEYIPGYTSK-----IFLPD-----DIKV-----SPLVKI 375
Db	1182	MIENMIENMIENMIENMIENMIENMIENMIENMIENMIENMIENMIENMIENMIEN 1236	Db 338 --ESWRIB-MGKYEQVDVDFDKFRSQVELNKNIYEDYKISQVDNIRREVELSLD 394
Qy	737	V---SANQYEVIRNEEGREILDHGKWKINKEESIKD-----ISSKEYISFNP 782	Qy 353 KSDKSE-----IFLPD-----DIKV-----SPLVKI 375
Db	1237	IEDDAKQINQVQNNIQNVHVKDKNDLNNVD--IINDVLKSDDKFENNINSKE---INI 1292	Db 395 LNSKMSVQSGAIDFKIREDSDNGIYLEFKPGKFCADIETVPSFSGKDINQLKMWLESQ 454
Qy	783	KENKIIVKSK-----YLHELSTLLOEIRNANSDDILEKKWMLTECEINVASNIDRQIVE 838	Qy 376 APANNVSINOALISLSDSYSCSDL---VINQIKRYKILNDNLN---PSINEGTFNTWT---427
Db	1293	KDFKVDKNDICLNLNELNK--QSYNSNTTEKCDIH-----DIFYENKQ---VH 1339	Db 455 LDVDNS-IQEKLIKILNDNLN1SNFEEINGRFNNYNSLNDNINAKYTALFESLSDSSSKPE 513
Qy	839	GRIEAKNLTSINSYIKNEFKLIESISLVDLKHQGLDSDSHFISPDISKTENGRI 898	Qy 428 -----MKIFSDKLASISNEDNMMEIKITNYLVGFPADVRSITNLSPGVYTGAYQDL 481
Db	1340	DKIHEDRKEYCD-----ESKLYLPADDIKMR-----SPKINKGKRENFLLI 1384	Db 514 NOMESKYSFTDKLTAGMDEFSLMYGEF-----ETUSQEA-----TNNYQB- 555
Qy	899	RFINKETG---NSIFIE-TEKEIFSEYATHISKE-ISNIKDTIFDNVNGKLVKYNLDA 953	Qy 482 LMFKDNSTNHLLEPELRNFFPKTKISQITEQIETSLWSFNQARAKSQPEYKKGVEG 541
Db	1385	SYMNEEYIPLIKKLVIEIPEISILNKNMHIKQFLSPGKYNDNLK-----1432	Db 556 --FQD-----LNKLENEIESFY-----NMFEKTQ-----578
Qy	954	HEVNTLNSAPFTQSLIENYNTTKESLN 980	Qy 542 ALGEDDNLDPQNTVLDKDYVSKKILSSMKTRNKKEYIHYIYVLOQGDKISY-----591
Db	1433	-EGKCISKIIIDLMLEYNTIISKNN 1458	Db 579 ---ETLKVDFTNTSLINIKDEIGKNIVEPRDRYDEVNIFVTQLEESKLYSKWQGMDSN 635
RESULT 11			Qy 592 ---EASCNLFKOPYSSILYQKN--IEGSETAYYYVADAEIKED-----632
G70163			Db 636 LKNIESQINKNEEFLSLIQOKGIGELSSVFNLDSDHIQKKAIDMHGSKWDELIANL 695
hypothetical protein B80512 - Lyme disease spirochete			Qy 633 KYRIPYQISNKNRIKLTIFGHGKSEFTDTFANLVDVDSLSSEI-----ETILNLAKADIS- 687
C:Species: Borrelia burgdorferi (Lyme disease spirochete)			Db 696 KSLDLKVSSELL-----SSATLKIESLEKDVNDRMENVL-LKTGDIES 739
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999			Qy 688 ---PKYIEINLGCNMFYSYIABETYPGKLLKIKRVSLEMPISIS-----QDSITVS- 738
C:Accession: G70163			Db 740 LVIEYKELKDMYSQSDEALIGIKFEINRQTEIITKDSVFMLEDLKKFKDDKNNFVISK 799
R:Praser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; White			Qy 739 -----ANQVEYVINE--EGKREI-----LDHSGKMI- 762
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Vugt,			Db 800 IEEDYKULQFKISEDILNFKSLDNLFBSKLIQVSNIKSDNOKQIDDFLDRIKDL 859
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horat, K.; Roberts, B.			Qy 763 NKEESIIDKSSKEYISFNPKENKIIKVSRYLHELSTLLQEIIRNNANSDDIL-----EK 817
Nature 390, 580-586, 1997			Db 860 NKEDSINNVEVSK-LSDQSKLNEITVK-----IENLLSGKVLDLIDSEV 905
A:Authors: Smith, H.O.; Venter, J.C.			Qy 818 KVMLTECEINVASNIDRQIVEGRIBEAKN-----LTSDSINYKNEFK-LIESIS-- 866
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.			Db 906 TTKIKELKFSIES-LESYVLE-KIDFPRNQGAISYDELLQDIDMHNFKETRELENLSKK 963
A:Reference number: A70100; MUID:98065943; PMID:9403685			
A:Accession: G70163			
A:Status: preliminary; nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-2166 <KLE>			
A:Cross-references: GB:AE001153; GB:AE000783; NID:G2688419; PIDN:AAC66876.1; PID:G268842			
A:Experimental source: strain B31			
Query Match			
Best Local Similarity			
6.1%; Score 318.5; DB 2; Length 2166;			
20.1%; Pred. No. 0.00017;			

```
QY 867 -----DSLYDLKH-----QNGLD-----DSHFISFEDISKTENGFRIRFIN 902
Db 964 PAAVLNNSBEFVKEVDSLLQDKETDIASQANIDITLDSLNVKFNDEINKEINGKYNEVIS 1023
QY 903 KETGNSIFITEKEIFSEYATHISKEISNIKOTIFONVNGKLVKKVNLDAAEHVNTLSA 962
Db 1024 NYRGYS-----ENISSKLENIEMHEIENLSRRLTDRIID-SLSKGM-----1063
QY 963 FFIQSLEIYNTKESLSNLVAMKVQVY-AQLFSTGLNTIID--ASKVVELV-----S 1012
Db 1064 -----ENLQKLKESPDVSKYQVERFELVKVOLDTDDGEAKINKLVKEIEQYK 1110
QY 1013 TALDETID 1020
Db 1111 SRLFEAID 1118

RESULT 12
G81702
adherence factor TC0437 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81702
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: G81702
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3255 <TET>
A:Cross-references: GB:AE002311; GB:AE002160; NID:g7190477; PID:AAF39291.1; PID:g719048
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0437

Query Match 6.1%; Score 316.5; DB 2; Length 3255;
Best Local Similarity 19.4%; Pred. No. 0.00035;
Matches 247; Conservative 197; Mismatches 472; Indels 359; Gaps 54;

QY 3 LVNAQLQKVVVVFRIQDEYVAILNALBEEVHNMSSESVVEKYLKDKDINNLTNYLNT 62
Db 200 IASRLKLTBEQIOKCHISLNN---LKAQRYLNIDESE--RNSRKGQDLLVKQSIFLNN 254
QY 63 YKSGRNKALKPKFYLTVWEVLKNNSLTPVEKNLHFIWGGQINDTAINYNQWQVY 122
Db 255 ILENAKWEESEEPKNVMSVIKTEF-NSHRVKIDSHPHGIWAGAPPEGTDVVIKTPLOTY 313
QY 123 SDYTVKFVVDNAPLNTLTKTIVESATNTLTLESFRENLDNDEFDYNYKEY-----172
Db 314 BEPDFLFWDRQAFGAAKFSSILKTIADFSSLKELRSITSQETQDFVKYDELOAKYQTS 373
QY 173 -----RKRMEIYKQKHID-----YKQSQIENPEP-----IID 203
Db 374 GRREERRELEKOLKFLDHYKLSQIRTFDAPLFLKNVLSQDGFNFYCLKGIGTIDD 433
QY 204 NIKTYSNEYSKDLBALNKY-----IEESLNKITANNND-----IRNLEKFA 247
Db 434 QTRIALQDELHPLEEIEIQYKDLIRANEQKIDIVGKVTDLGKERVEIKNIRDLSMQ 493
QY 248 DEDVLRYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKSINK-----303
Db 494 DRTNTYNEEMELLRWYPAASQIRMYMLKEUGGIYTDLDMMFPQYSPDVLQMINBIGD 553
QY 304 -----303
Db 554 RPEELPIRRAVSDGVRLANGESGITIDNIAQDIDISKLTRSDRTQLEKLLTDIENKQ 613
QY 304 -----PSGIT-----NTSEWMIKLEALIMKYKEYIPGVTYSKNFDM 337
Db 614 SSGTSKFSLFQRMADPSVRDFMFIQRRHKKWSTGMVNRGLNGLM-----MAHKGSA 665
```

```
QY 338 LDEEV--QRSFESALSSKSD--KSEIFLPLDDIKVSPLEVKIA-----F 377
Db 666 VDAVIGKQRAYRELKSLRETIVLSGEFFKTLDDLKHNLHAKLIGGHLVEDFLGKSLFSEF 725
QY 378 ANNSVINQALISLKDYSYCDLVINQIKNRYKILNDNLNPSINEGTFDNTTKWIFSDKLAS 437
Db 736 RQDTIAGALSTLGIT-GPDLIVDEMKTYSFRSLGP-IGKDVLEGKKLGN--KAPLGSYNK 781
QY 438 ISNEDNMWF--MIKITNVLKVGFPADVRSTINLSGPGVYTGAYQDOLLKFKONSTIHLE 495
Db 782 IVKEGKQTFDMLHPVT---VG-ANDVTPADESTWCGIKQRCIAELL--FSDESK---FR 831
QY 496 PELRNFPEPK---TKISQLTEQEITSLWSFNQARAKS-----QPEEYKKGIFEGALGED 546
Db 832 PET-----PKGISRTKIDKKT---FTKLWSEOSKALSPELLDRFNE-----LIEA 874
QY 547 DNLDFAQNTVLDKO-YVSKKILSSMKTRNKYIHYIVQ---LOGDKISYEASCNLF---598
Db 875 KNFDLTKFTDVRDALYVLRNQVCCDASAAVFSIQLQLAELLRCTSPPIANQLNVFPDL 934
QY 599 SKDPYSSI-----LYQKNIEGSETAYY-----YVADAE---IKE--IDKVR 635
Db 935 NRNIASSLEKTIKLYLOSHSOTETVIHSAISDRVMFLRDLMLATAERMLVKESIFSKEE 994
QY 636 IPYOISNKRNIKLTFIGHGKSEFNTDTFANLDVDSLSSEIEFILNLAKADISPKVIEINL 695
Db 995 TP--UTTEIKLNSWAEQLSKDNUGILSBEVDKLDVTSIMENSKLKKRVSALDSI 1052
QY 696 LGCNMFYSIYAET-----YPGKLLLIKIDRVSELMPISISQDSITVSANOYEVRINBE 749
Db 1053 VAGYFY---LRLEETLSEWVRIPPKDLKRVIRFMEGMDANQHPDERGGRWYSQLYDES 1109
QY 750 GK-----REILDHSGWINKBESIIKDSSKEYISFNPKENKIIIVKSKYLHELSTLLQ 802
Db 1110 FKXRVTPAPVKKIQLAKKYLNEQRVHVLEID--DYLTKNPLFNR-----LHE-----1154
QY 803 EIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGRIEAKNLTSDSI-----NYI 855
Db 1155 ---EGYAFSDUTEITRYMLAEGIS-----GIFSEGNILPSPSARLVNIKITV 1200
QY 856 KNEFFKLIESISDSLVDLKHQNG---LDDSHFISFED-----ISKTEN 894
Db 1201 GGDYHDMQDVLPIKYDWLASGAADLTNERPAAIPESLRKNLEGLHGTDLTLPVDASVS 1260
QY 895 GPRIRFINKEGNSIFITEKEIFS---EYATHISKEISNIKDTIFD-----NVNGK 943
Db 1261 GWMGSF-GVENG---VESDHTMISIAFGFFNGASYSQHYLSALYEIHRHIHLSLTSD 1315
QY 944 LVKVKVNLDAAEHVNTLSNAFFI---QSLIEYNTTKE--SLSNLSVAMKVQVY-AQLFS 995
Db 1316 LIKK-----ELESKGAGCFVHEERFDSLLKASSEKQVLSLUTEIHKLSNOVHLAEVS 1368
QY 996 TGLNT-ITDASKVVE 1009
Db 1369 HLMNTALPGVGKIE 1383
```

## RESULT 13

```
Ti8501
hypotheetical protein C0760c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: Ti8501
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: Ti8501
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3394 <LAW>
A:Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1
C:Genetics:
```

A:Map position: 3  
A:Note: C0760C

```

Query Match          6.1%; Score 315.5; DB 2; Length 3394;
Best Local Similarity 18.7%; Pred. No. 0.0004;
Matches 247; Conservative 212; Mismatches 422; Indels 439; Gaps 57;

QY 1 MNLVNAQLQKVVYKFRIOEDYVAIL-----NALEEYHNMSESSVEKYLKLDINNLL 55
DB 1722 MWILKXKETENM--NIQKEDYIKLIKDKKTNIQNEYNDL-----LEKYNEVVVKNNM 1773
QY 56 TDNYLNTYKSGRNKALKKPKYL-----TMEVLEKNNSLTPVSKN-LHFTWIGQIND 109
DB 1774 LYNDNV-----LLKEHEEIEFLKENIKILOKNTYLMDFKQINNV-----DNN 1820
QY 110 TAINVINQWKNVSP-----YTVKVFY 131
DB 1821 LUKRLDQFLNQLDQKHLDTNQKHLQKVDYIEIKERLKIETKINKQEKYIILOK 1880
QY 132 DSNAPLINTLKTIVESATNTLTESFRENLDNPEFDYKFY-----172
DB 1881 DNN-LILNDFNSTTTTNNNNNNNNNNNNNNNNNNNNNDNDYQOFIHSKANLENSRLK 1939
QY 173 -----RKRMEIYDQ-KHFIDYKQ-----IENPEFIID-NIITYL- 210
DB 1940 ELSNLNEKIQLSDEKNRKITILEDKLPKNEKDKMKLQIIDDNNKNYMIQYKLNKLTLD 1999
QY 211 -----SNEYSKOLEALNKYIEESLKNITANNNDIRNLEKADE-----249
DB 2000 MLSEENRMULLNKEYEKQIEQLNH--DHKLPISTNNNDIQIIEENEKLOEQVDYITIN 2057
QY 250 -----DLVLYNQELVER-----WNLAAASDILRIS--MLKED--GGVLDVLD- 288
DB 2058 EKDKIIVHLNQIKLANQNEHRSRCDIFNVAHQSDNIKNHVMVGEDINGDTHDYNK 2117
QY 289 -----ILPGIQDPLFKSINKPSINTNTSWEMIKLEAIM-----KYKE-----YIPG 329
DB 2118 NIDQGTNQHINOQTNQHOQTDTCGPNYVYVQVQVQVQVQVQVQVQVQVQVQVQVQV 2177
QY 330 YTSKPFMDLDE-----EVQSFESALSSKSDSEIFLPLDDI-----KVSPL--- 371
DB 2178 YINENIDLTSELEKNDMLNENKELKEKNE--EYKLNNDIDMLSNCKKLESGIMME 2235
QY 372 EVKIAFANN-----SVINQALIS-LKDS-----YCSDL- 398
DB 2236 KYKIIIMNNIOEKDEIENLKNYKNNKLLDNLNYSVDKSVSCFEDSNIMSPSCNDIL 2295
QY 399 -VINOIKNRYK-----ILNDNLN--PSINEGDTFNTTKIFS-DKLASISNEDNMFM 447
DB 2296 NVFNLSKSNKKVCTNMDCENMDSISSINNWNINNWNINNWNINNWNINNWNINNWN 2355
QY 448 IKITNLYKVGAPDVRSTINLSGPGYVTCAYQDILLMPKONSTNIHLPELBNPEPKTK 507
DB 2356 VDINYLNNL-----QLNKDNDIIIIKFNILKFLGSCYLYIIRNLKEIQMLKNQ 2409
QY 508 ISOLTEOEITSLWSP--NOARAKSOFEEYKGYFEGALGEDNLOFAQNTVLD-KDYVSK 564
DB 2410 ILSL-BESIKSLNEFINLNKNEKNEKELIKINNFEILKLNKLNQDNESCIQNLNLYLK 2468
QY 565 -KILSMKTRN-----KEYTHIVQLQGDKISYEASCNLFSKDPYSILYQKNIEG---SE 616
DB 2469 NEELNKNVKNIPKYGYIHLIQ-----QSNVFCIKPFKNENKIDQSIINKLLYIK 2522
QY 617 TAYYVYVADABEIKDKYRIPQISNKNRIKUTFTGHGKSEFNTDTFANLDVDSLSSEIE 676
DB 2523 KSFDFMYDVSIOE-----IRENKNIII-----NODFLTDEY-----FKHIG 2559
QY 677 TILNLAKADISPKYIEI-----NLLG-----CNMFSYS-----704
DB 2560 TPTKTCNVLIQRYGISILKDTNDFIYNKQSQNGQNGNHNMCNIIYPPDEINVTADQ 2619
QY 705 -----IYAEETYPGKLL-----717

```

```

DB 2620 QIFDGTENVQOSLQNEEDYVNVNEEMYTDKMDLNNMNGDDDDDDDDDDNNNNNNNN 2679
QY 718 -----KIKDRVSELMPSISQDSITVSANOYEVRIINECKREILDHSGKWINKESIIKDI 772
DB 2680 NNNNNMGDDENHLVAFNNHLLTNGNVKSDQINNETLRY-----EENIQNI 2729
QY 773 SSKEYISFN--PKENKIIIVKSY-----LHELSTLLQOIRNANSSDIDLEKK 818
DB 2730 YTNDDVNDQVNIENKILIKQDQIINNDELKNEHNNLIRLINESIENAH-----NLENV 2785
QY 819 VMLTECEINVASINDROIVEGRIBEAKNLTSDS-----INVIKNEFKLIESI 865
DB 2786 YVQNDANNLINDNTKKESETLTYVDEKONVSNEKSCDDDKKKNBDIIQAKNENYFVSTH 2845
QY 866 SDSLYDLKHQGLDSDHSPISFEDISKTEGFRIRFINKETGNSIFIETEKEIFSEVATHI 925
DB 2846 YDNDDINKNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNN 2892
QY 926 SKEISNIKDTTIDFNVNGKLV-----KKVNLDAHAHEVNTLSAFAFIQSLIEYN 972
DB 2893 NINDDNNNDNNNDNNNGFVCELSSNINDFNNILNVKDNFGINKSN--FSTNLSEYN 2950

RESULT 14
JC6009
surface-located membrane protein lmp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JC6009
R:Jadefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system
A:Reference number: JC6009; MUID:96213016; PMID:8631664
A:Accession: JC6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197336
C:Genetics:
A:Gene: lmp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>
F:957-992/Domain: tetratricopeptide repeat homology <T1>
F:993-1026/Domain: tetratricopeptide repeat homology <T2>
F:1089-1120/Domain: tetratricopeptide repeat homology <T3>
F:1154-1190/Domain: tetratricopeptide repeat homology <T4>

Query Match          6.0%; Score 311; DB 1; Length 1302;
Best Local Similarity 20.1%; Pred. No. 0.00017;
Matches 234; Conservative 204; Mismatches 396; Indels 330; Gaps 53;

QY 3 LVNKAQLQKVVYK-FRIOEDYVAILNALSEYH-----NMSESSVVE--- 44
DB 169 LLNKINAEERLQSKIFNEKKQELKRVLD-LEDTEVDFTKEQVFIETINETSIEDIK 227
QY 45 -KYLKL-KDINLTDNYLNTYKSGRNKALKKPKYLMEVLEKNNSLTPVSKNHLFTW 102
DB 228 NKIIIEVERKATSLTSKIINT-----KQELQEF-ENIKKLDQDFINTKLNDAKYQ----- 276
QY 103 IGGQINDTAIVINQWKNVSDYTVKFPVYVDSNAFLINTLTKTIVESAT---NNTLESPE 159
DB 277 ---SLKQALDKINSNGINKNSTIKEKAGONALIKAKEEAGLEKEKLDGQNIKOTLKE 333
QY 160 NLNDPEFDYKFRKRMEEIYDKQKHFDYKYSQIEENPEFIIDNIKITLSNEYSKOLE 219
DB 334 TINNAK-----EPKLLIDNDQKIVD--LKSNDLONEISKARQ 368
QY 220 ALNKYIESLNKITANNNDIRNLEKFADEDLRLYLNQELVERWNLAASDILRISMLKE 279
DB 369 SLSK-----DKMESANDLLNTKLBYKELNFKNOEKAFFN-----ELEQTRK 414

```







Thu Nov 6 08:28:34 2003

us-09-126-816b-6\_copy\_1\_1020.rag

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:02 ; Search time 30.75 Seconds  
(without alignments)  
5265.077 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1020

Perfect score: 5192

Sequence: 1 MNLVNAQLQKVVYKFRIQ.....ITDASKVELVSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4086	78.7	2366	17 AAR95011	C. difficile toxin
2	4083	78.6	19 AAW68388		Clostridium diffic
3	2783	53.6	2710 17 AAR95016		C. difficile toxin
4	2783	53.6	2710 19 AAW68387		Clostridium diffic
5	2780	53.5	546 20 AAV33700		C. sordellii letha
6	293.5	5.7	1254 11 AAR07503		Merozoite apical-en
7	293.5	5.7	1254 18 AAW24575		Merozoite apical-e
8	286.5	5.5	1979 21 AAB18171		Plasmodium falcipa
9	280	5.4	1516 21 AAB18195		Plasmodium falcipa

10	272	5.2	980	21	AAB18294	Plasmodium falcipa
11	266.5	5.1	1121	21	AAB18241	Plasmodium falcipa
12	261.5	5.0	2184	22	AAE00425	P. falciparum telo
13	258.5	5.0	1712	21	AAB18205	Plasmodium falcipa
14	256	4.9	1090	23	ABB54799	Lactococcus lactis
15	256	4.9	1948	23	ABP73774	Candida albicans e
16	255	4.9	1817	21	AAB18301	Plasmodium falcipa
17	255	4.9	2485	21	AAB18172	Plasmodium falcipa
18	249	4.8	1639	19	AAW54145	P. falciparum synt
19	249	4.8	1639	23	AAE29345	Plasmodium falcipa
20	247.5	4.8	10182	23	ABP38314	Staphylococcus epi
21	246	4.7	1881	23	ABP73809	Candida albicans e
22	245.5	4.7	5024	22	AAE82935	S. epidermidis ope
23	242.5	4.7	1346	21	AAB18236	Plasmodium falcipa
24	241.5	4.7	1089	23	ABB09491	Ampev first DNA po
25	238	4.6	1477	20	AAV19981	B. burgdorferi ant
26	238	4.6	1494	20	AAV19980	B. burgdorferi ant
27	237.5	4.6	3433	18	AAW22017	Utraphin. Homo sa
28	237	4.6	1654	6	AAE50777	Sequence of the p1
29	236.5	4.6	1308	21	AAB18167	Plasmodium falcipa
30	236.5	4.6	1802	21	AAB18217	Plasmodium falcipa
31	235	4.5	3210	24	ABU07438	Protein differenti
32	235	4.5	3248	17	AAE99795	Kinetochore protei
33	232	4.5	3973	21	AAB18253	Plasmodium falcipa
34	229.5	4.4	2681	24	ABJ19025	Pathogen specific
35	224.5	4.3	1151	23	ABP38397	Staphylococcus epi
36	224	4.3	1270	23	ABP73290	Candida albicans e
37	222.5	4.3	2663	22	AAE39097	Human polypeptide
38	221.5	4.3	1558	21	AAB18324	Plasmodium falcipa
39	220.5	4.2	2295	21	AAB18180	Plasmodium falcipa
40	220	4.2	1188	21	AAB18183	Plasmodium falcipa
41	220	4.2	3696	23	ABP40235	Staphylococcus epi
42	219	4.2	1714	21	AAB18275	Plasmodium falcipa
43	219	4.2	2496	21	AAB18222	Plasmodium falcipa
44	218.5	4.2	1145	22	AAE82169	S. epidermidis ope
45	218.5	4.2	1430	20	AAV31949	Plasmodium falcipa

#### ALIGNMENTS

RESULT 1  
AAR95011  
ID AAR95011 standard; Protein; 2366 AA.  
XX  
AC AAR95011;  
XX  
XX  
DT 08-JUL-1996 (first entry)  
XX  
DE C. difficile toxin B.  
XX  
KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;  
XX diarrhoea; therapy; diagnosis; vaccine.  
XX  
OS Clostridium difficile VPI strain 10463 (ATCC 10463).  
XX  
XX  
XX WO9612802-A1.  
XX  
PD 02-MAY-1996.  
XX  
PF 23-OCT-1995; 95WO-US13737.  
XX  
PR 07-JUN-1995; 95US-0480604.  
PR 24-OCT-1994; 94US-0329154.  
PR 16-MAR-1995; 95US-0405496.  
PR 14-APR-1995; 95US-0422711.  
PA (OPHI-) OPHIDIAN PHARM INC.  
XX  
PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
PI Williams JA;  
XX  
DR WPI; 1996-230603/23.

DR N-PSDB; AAT29247.  
XX Fusion proteins comprising non-toxin protein and part of toxin -  
PT useful to form anti-toxins against Clostridium botulinum type A, and  
PT C. difficile type toxins, and to treat C. difficile intoxication,  
PT partic. diarrhoea  
XX  
PS Claim 36; Page 313-323; 434pp; English.  
XX  
CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the  
CC product of the toxin B gene (AAT29247), is a cytotoxin associated  
CC with diarrhoeic disease. It can be obt'd. by expression in  
CC transformed E. coli hosts of portions of DNA that together cover the  
CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,  
CC AAR75371-72 and AAR95018), pref. expressed as fusions to polyhistidine  
CC affinity tags or maltose binding protein, are used to raise avian  
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.  
XX  
SQ Sequence 2366 AA;  
Query Match 78.7%; Score 4086; DB 17; Length 2366;  
Best Local Similarity 77.4%; Pred. No. 1.5e-213;  
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;  
QY 1 MSLVNRKALQKQVYVFRQEDYVAIINALEYHNHNSSESVVEKYLKLDINNLTDNYL 60  
DB 1 MSLVNRKQLEKMANVFRQEDYVAIINALEYHNHNSSESVVEKYLKLDINSLTDYI 60  
QY 61 NTYKSGRNKALKKFKYELTMEVLELKNLSLTPVKNLHPITWGQINDTAINYNQWKD 120  
DB 61 DTYKSGRNKALKKFKYELTMEVLELKNLSLTPVKNLHPITWGQINDTAINYNQWKD 120  
QY 121 VNSDYTVKPYVDSNAFLINTLTKTIVESATNTLESFRENLANDEPDYKFKRMEIY 180  
DB 121 VNSDYTVVNFYDSNAFLINTLTKTIVESATNTLESFRENLANDEPDYKFKRMEIY 180  
QY 181 DKQKHIDYKSGQIENPEFIIDNIKTLYNSYKDLALNKYIEESLNKITTANNNDI 240  
DB 181 DKQKNFVYKAQREENPELLIDIVKTYLSNEYSKEIDELNTYIEESLNKITTQNSNDV 240  
QY 241 RNLEKFADELRLVYNQELVERNLAAASDILRISMLKEDGGVYLDVLPQIPDLFKS 300  
DB 241 RNLEKFADELRLVYNQELVERNLAAASDILRISMLKEDGGVYLDVLPQIPDLFKS 300  
QY 301 INKPDSTNTSWEMIKLEAMTKYKEVTPGYTSKNFDMLEEVQSPESALSSKSEIF 360  
DB 301 IEKPSVTVDFWEMTKLEAMTKYKEVTPGYTSKNFDMLEEVQSPESALSSKSEIF 360  
QY 361 LPDLDIKVSPLEVKIAFANNSVINQALISLKQSYCSDLVINQIKNRYKIINDNLPINE 420  
DB 361 SSGDMEASPLEVKIAFNSKGIINQGLISVKQSYCSNLIVKQLENYKILNLSNPAISE 420  
QY 421 GTDFNTMTKIFSKLASINEDNMFMKITYNLKVGFPADVRSTNLGSGPVYTGAYQD 480  
DB 421 DNDFTNTTTFIDSIAMAEANADNGRFMWELGKLVGRVFPDVKTNTNLGPEAYAAAYQD 480  
QY 481 LLMFKDNSTNHLLEPLRNFPKTKI SOLTEBITSLWSFNQAKSQFEYKKGYPE 540  
DB 481 LLMFKGSMNHLIEADRLNFEISKNTI SOSTEQEMASLWSFDDAKAQFEYKKNRYFE 540  
QY 541 GALGEDNDLDFQNTVLDKDYVSKKILSSMKTKNKEYIHYIVOLQGDKISYEASCNLFQK 600  
DB 541 GSGLEDNDLDFSNIVVDKEYLLEKISLARSSERGIHYIVOLQGDKISYEACNLFQK 600  
QY 601 DPVSSILYQKNIGSETAYVYVADAEIKEDIKRIPYQISNKRNTKLTFIGHGKSEFNT 660  
DB 601 TPVDSVLQKNIEDSEIAYVYVADAEIKEDIKRIPYQISNKRNTKLTFIGHGKSEFNT 660  
QY 661 DTANLNDVLSLSEIETILNLAADISPKYIEINLGCNMFSYIYAETYPCKLILKIK 720  
DB 661 DIFAGFDVSLSTEIAEDLAKEDISPKYIEINLGCNMFSYIYAETYPCKLILKIK 720  
QY 721 DRVSELMPSISQDSITVSANQYEVRIINEGKEILDHSGKWNKESIIKDISKEYISF 780

DB 721 DKISELMPSISQDSITVSANQYEVRIINEGKEILDHSGKWNKESIIKDISKEYISF 780  
QY 781 NPKENKIIVKSKYLHELSTLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840  
DB 781 NPKENKIIVKSKYLHELSTLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840  
QY 841 IEENKALTSDSINYIKNBFKLTIESDSLYDLKHONGLDSDHFI SPEDISKTENGPRIRF 900  
DB 841 IEENKALTSDSINYIKNBFKLTIESDSLYDLKHONGLDSDHFI SPEDISKTENGPRIRF 900  
QY 901 INKETGNSIFIEETEKEIFESEYATHISKEISNIKDTIFONVNGKLVKKNLDAHEVNTLN 960  
DB 901 INKETGESIFVETEXTIFSEYANHITERLSKIGTIFDVNGKLVKKNLDTTHEVNTLN 960  
QY 961 SAFFQTSQSLIYNTTESISNLSNVAMKVQVYAOQFSTGLNTIITDASKVELVSTALDETID 1020  
DB 961 SAFFQTSQSLIYNTTESISNLSNVAMKVQVYAOQFSTGLNTIITDASKVELVSTALDETID 1020  
RESULT 2  
AAW68388  
ID AAW68388 standard; Protein; 2366 AA.  
XX  
XX AAW68388;  
DT 07-DEC-1998 (first entry)  
XX  
DE Clostridium difficile toxin B.  
XX  
XX Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;  
KW pseudomembranous enterocolitis.  
XX  
OS Clostridium difficile.  
XX  
PN W09808540-A1.  
XX  
PD 05-MAR-1998.  
XX  
PF 28-AUG-1997; 97WO-US15394.  
XX  
PR 28-AUG-1996; 96US-0704159.  
XX  
PA (OPHI-) OPHIDIAN PHARM INC.  
XX  
PI Thalley BS, Williams JA;  
XX  
DR WPI; 1998-230234/20.  
DR N-PSDB; AAV30561.  
XX  
PT Host cell containing recombinant expression vector encoding  
PT Clostridium botulinum type B or E toxin - useful to treat humans  
PT and other animals at risk of intoxication with clostridial toxin  
XX  
PS Example 18; Page 241-249; 428pp; English.  
XX  
CC This is the amino acid sequence of Clostridium difficile toxin B,  
CC deduced from the coding region (see AAV30561) of the toxin B gene.  
CC Fragments of the toxin B gene have been cloned into various  
CC prokaryotic expression systems, and assessed for the ability to  
CC express recombinant toxin B protein in E. coli. It would be  
CC advantageous to use simple and inexpensive prokaryotic expression  
CC systems to produce and purify high levels of recombinant toxin B  
CC for immunisation purposes. The invention specifically relates to  
CC recombinant proteins derived from Clostridium botulinum toxins  
CC (see AAW68389-400) and their use as immunogens for the production of  
CC vaccines and antitoxins.  
XX  
SQ Sequence 2366 AA;  
Query Match 78.6%; Score 4083; DB 19; Length 2366;  
Best Local Similarity 77.4%; Pred. No. 2.1e-213;  
Matches 789; Conservative 106; Mismatches 125; Indels 0; Gaps 0;



QY	1	MNLVNAQLOQMYYVKFRIOEDYYVAIINALBEYHNMSSESVVEKYKLKDINNLTNYL	60
DB	1	MSLVNRKQLEKANVRFRIOEDYYVAIINALBEYHNMSSESVVEKYKLKDINSLTDIYI	60
QY	61	NTYKSGRNKALKKPKLEYLTMBVELKNNSLTPVEKNLHFIIWIGQINDTAINYQWKD	120
DB	61	DTYKSGRNKALKKPKLEYLVTEVLEKNNLTPVEKNLHFVWIGQINDTAINYQWKD	120
QY	121	VNSDYTVKPVYDSNAFLINTLAKTTVESATNTLTSPRENLDNPPDYNNKPKYKMEIY	180
DB	121	VNSDYNNVYFDSNAFLINTLAKTTVESAINDTLTSPRENLDNPPDYNNKPKYKMEIY	180
QY	181	DKQKHFIDYKQIOENPEFIIDNIITKYLNSYKDLKLEALNKYIETESLNKITTANNNDI	240
DB	181	DKQKHFINYKKAQREENPELLIDIVKTYLSNEYSKEIDELNTYIETESLNKITSQSGNDV	240
QY	241	RNLEKPADBDLVRNLQNELVERWNLAASDIILIRISMLKEDGGVYLDVILPGIQDLPKS	300
DB	241	RNFEEFKNGESFNLYEQELVERWNLAASDIILIRISALKEIGWYLDVMDLPGIQDLPFS	300
QY	301	INKPDSITNTSEMILKLEIMKYKEYIPGYTSKNFDMLEEVQORSFESALSSKSKSEIF	360
DB	301	IEKPSGVTVDFWEMTKLEAIMKYKEYIPEYTSHEFDMLEEVQSFESVLASKSKSEIF	360
QY	361	LPDLDDIKVSPLEVKIAPANNSTVINOALISLKDSYCSDLVINOIKNRYKILNDNLPSINE	420
DB	361	SSLDMEASPLEVKIAPNKGKIINOGLISVKDSYCSNLIVKQIENRYKILNNSLNPASE	420
QY	421	GTDFNTMTKIFSDKLASISNEDNMFMKITYLYKGFAPDVRSTINLSGPGVYTGAYOD	480
DB	421	DNDFNTMTTFIDSIAMAEANADNGRPMELGKYLVRGFPDPDKVTINLSGPEAYAAAYOD	480
QY	481	LMFKNSTNIHLPELBNFPPTKISQLTQEBITSLWSFNQARAKQFPEYKKGYPE	540
DB	481	LMFKEGSMNIHLIEADLNFEISKTNIISQTEQEMASLWSFDDARAKAQFEYKRYEPE	540
QY	541	GALGEDNDLDFACNTVLDKDYVSKITLSMKTRNKEYIHVIVLOQDKISYEASCNLPK	600
DB	541	GSLEGDDNDJFQNIIVVDKEYLEKISSLARSGERYIHVIVLOQDKISYEAACNLPK	600
QY	601	DPYSILYQKNEGSETAYYYVADAEIKEDKIRIPIYQISNKRNIKLTFIGHGSEFNT	660
DB	601	TPYDSVLFPKNIEDSEIAYVYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNT	660
QY	661	DFANLDVDSLSSEIETILNLAADISPKYIEINLLGCMFYSIYABEITYPKLLLKIK	720
DB	661	DIFAGFDVDSLSTETBAADLAKEDISPKSIIEINLLGCMFYSINVEETYPCKLLLVK	720
QY	721	DRVSELMPISISODSITVSANQYEVRIINEGKREILDHSGKWINKEBSIIKDSSKEYISF	780
DB	721	DKISELMPISISQDSIIIVSANQYEVRIINSGRELLDHSGEWINKEBSIIKDSSKEYISF	780
QY	781	NPKENKIIYKSKYLHELSTLLOEIRNANSSDIDLEKVMLTCEINVASNIDRQIVEGR	840
DB	781	NPKENKITVKSKNLPELSTLLOEIRNNSSDIELEKVMLTCEINVISNIDTQIVEER	840
QY	841	IEBAKNLTSDSINYIKNEFKLTESISDSLYDLKHONGLDSDHPISEPIDSKTENGPRIRP	900
DB	841	IEBAKNLTSDSINYIKDEFKLTESISDALCDLQKQNELEDSHPISPESIDSETDEGFSIRP	900
QY	901	INKETGNSPIFTEKEIFSEYATHISKEISNTIKDTIFDNVNGKLVKKNVLDAAHEVNTLN	960
DB	901	INKETGESIPVTEKTFISEYANHITETESKIKGTIFDTVNGKLVKKNVLDTHTEVNTLN	960
QY	961	SAFFIQSLIEYNTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELSVSTALDETID	1020
DB	961	AAFFIQSLIEYNSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVVELSVSTALDETID	1020
RESULT 3			
AAR95016			
ID	AAR95016	standard; Protein; 2710 AA.	

XX	AAR95016;		
XX	AC		
DT	08-JUL-1996	(first entry)	
XX	C. difficile toxin A.		
DE	Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;		
KW	vaccine; diarrhoea; diagnosis; therapy.		
KW	Clostridium difficile VPI strain 10463 (ATCC 10463).		
XX	OS		
PN	WO9612802-A1.		
XX	02-MAY-1996.		
XX	23-OCT-1995;	95WO-US13737.	
XX	07-JUN-1995;	95US-0480604.	
PR	24-OCT-1994;	94US-0329154.	
PR	16-MAR-1995;	95US-0405496.	
PR	14-APR-1995;	95US-0422711.	
XX	(OPHI-) OPHIDIAN PHARM INC.		
XX	Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;		
PI	Williams JA;		
XX	WPI; 1996-230603/23.		
DR	N-PSDB; AAT29248.		
XX	Fusion proteins comprising non-toxin protein and part of toxin -		
PT	useful to form anti-toxins against Clostridium botulinum type A, and		
PT	C. difficile type toxins, and to treat C. difficile intoxication,		
PT	partic. diarrhoea		
XX	Claim 63; Page 290-302; 434pp; English.		
PS	Clostridium difficile VPI strain 10463 toxin A (AAR95016), the		
XX	product of the toxin A gene (AAT29248), is a potent cytotoxin that		
CC	plays a direct role in damaging gastrointestinal tissues and is		
CC	associated with diarrhoeic disease. It can be obtd. by expression in		
CC	transformed E. coli hosts of portions of DNA that together cover the		
CC	entire toxin A gene. Toxin A, and portions of it (see also		
CC	AAR95014-15 and AAR95017), pref. expressed as fusions to polyhistidine		
CC	affinity tags or maltose binding protein, are used to raise avian		
CC	antibodies useful as antitoxins or diagnostics, and in vaccine prodn.		
CC	Sequence 2710 AA;		
SQ	Query Match 53.6%; Score 2783; DB 17; Length 2710;		
	Best Local Similarity 52.1%; Pred. No. 1.4e-142; Indels 8; Gaps 5;		
	Matches 534; Conservative 218; Mismatches 265;		
QY	1	MNLVNAQLOQMYYVKFRIOEDYYVAIINALBEYHNMSSESVVEKYKLKDINNLTNYL	60
DB	1	MSLSEKSEILKAY-SIRPRENEYKTLTNLDEYKLTNNNNENKYQLKKNESIDVFM	59
QY	61	NTYKSGRNKALKKPKLEYLTMBVELKNNSLTPVEKNLHFIIWIGQINDTAINYQWKD	120
DB	60	NKYKTSRRNALSNLKKOILKEVILIKNSNTSPVEKNLHFVWIGGEVSIDALEYIKQWAD	119
QY	121	VNSDYTVKPVYDSNAFLINTLAKTTVESATNTLTSPRENLDNPPDYNNKPKYKMEIY	180
DB	120	INAEYNIKLYWDSAEFLVNTLKAIVSSSTTBALQLEEBEQNPQDNMKFYKGRMEFY	179
QY	181	DKQKHFIDYKQIOENPEFIIDNIITKYLNSYKDLKLEALNKYIETESLNKITTANNNDI	240
DB	180	DRQKHFINYKSKQINKPTVPTTIDDIKSHLVSEYRDETVLESYRTNSLRKNSHGIDI	239
QY	241	RNLEKPADBDLVRNLQNELVERWNLAASDIILIRISMLKEDGGVYLDVILPGIQDLPKS	300
DB	240	RANSLFTEQELNLYSQELLNRLNLAASDIVRLALAKNFGVGYLDVMDLPGIHSDFKT	299

QY 301 INKPDSTWTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVORSFESALSKSKSEIF 360  
DB 300 ISRPSSIGLDRWEMIKLEAIMKYKEYINNTYSNFDKLDQQLKDNFKLIESKSEIF 359  
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSQSDVLINQIKNRYKILNDNLNPSINE 420  
DB 360 SKLENLVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVGNRYQFLNQHLPAS 419  
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITYLVKVGFPADVRSTINLSGPGVYTGAYOD 480  
DB 420 DNNFTDTTKIFHDSLFNSATAENSMTLTIAPYLVQGFMPPEARSTISLSPGAYASAYD 479  
QY 481 LLMPKDNSTNIHLEPELNFEPKTKISOLTEOETISLWSFNQARAKSOFEEYKGYE 540  
DB 480 FINLOENTTEKTLKASDLTEFPENNLSQLTEQEINLSWSPDQASAKYQFEKYVRDYG 539  
QY 541 GALGEDNDLDFAOITVLDKDY-VSKKILSS--MKTRNKYIHYIVOLQGDKISYEASCNL 597  
DB 540 GSLSSENGVDNFKNATLKNLNNKIPSNVVEAGSKYVHYIQLQGGDISYEATCNL 599  
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEDKYPVPQVQISNKNKIKLTFPIGHGK 655  
DB 600 FSKNPKNSIIQRM--NESAKSYFLSDGSEILELNKYRIPERLKNKSKVKVTFPIGHGK 657  
QY 656 SEFNTDTPANLVDLSLSEIETILNAKADISPKYIEINLGCNMFESYIYAEETYPGKL 715  
DB 658 DEFNTSEFARLSVDSLSEISSEFDTIKLDISPNVFNLLGCNMFSDYFNVEETYPGKL 717  
QY 716 LLKIKDRVSLMPSISODSTVSAQVYEVRAINSEGRKILDHSGKWINKKEESIIDKISSK 775  
DB 718 LLSIMDKITSTLDPVKNKSTIGANQYEVRAINSEGRKELLASHGKWINKKEEALMSDLSK 777  
QY 776 EYTSFNPKNENKIVSKYHELSTLQEIIRNNANSDDIDLEKKVMLTECEINVASNDRQ 835  
DB 778 EYIFSDIDNKLAKSKNIPGLASISEDIKTLILDASVSPDTKFIILNKLNISSIGDY 837  
QY 836 IVEGRTEEAENLTSNTYINKFELIESISDSLYLKHONGLDDSHFISFEDISKTENG 895  
DB 838 IYKELEPVKNIIHNSIDDDIDFENLLENSVDSELYELKLNLUDEKYLISFEDISKNNST 897  
QY 896 FRIRFINKTGNFIFETEKEIESEYATHISKEISNIKOTIFDNVANGKLVKKNLDAHE 955  
DB 898 YSVRFINKNGESVYVETEKEIEFSKYSEHITKEISTIKNSIITDVGNLLDNIQLDHTSQ 957  
QY 956 VNTLNAFFIQSLIEYNTKESUSLSVAMKVQVYQLFSTGLNTTIDASKVVELVSTAL 1015  
DB 958 VNTLNAFFIQSLIDYSSNKNVDNLSTSVKVLQYQLFSTGLNTIYDSTQLVNLISNAV 1017  
QY 1016 DETID 1020  
DB 1018 NDTIN 1022

## RESULT 4

AAW68387 standard; Protein; 2710 AA.

AAW68387;

07-DEC-1998 (first entry)

Clostridium difficile toxin A.

Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;

pseudomembranous enterocolitis.

Clostridium difficile.

W09808540-A1.

05-MAR-1998.

PF 28-AUG-1997; 97WO-US15394.  
XX 28-AUG-1996; 96US-0704159.  
PR (OPH-) OPHIDIAN PHARM INC.  
XX Thalley BS, Williams JA;  
XX WPI: 1998-230234/20.  
DR N-PSDB; AAV30560.  
XX Host cell containing recombinant expression vector encoding  
PT Clostridium botulinum type B or E toxin - useful to treat humans  
PT and other animals at risk of intoxication with clostridial toxin  
XX Example 15; Page 220-230; 428pp; English.  
XX This is the amino acid sequence of Clostridium difficile toxin A,  
CC deduced from the coding region (see AAV30560) of the toxin A gene.  
CC Toxin A is a potent cytotoxin that plays a direct role in damaging  
CC gastrointestinal tissues. Severe cases of C. difficile  
CC intoxication result in pseudomembranous colitis. This would be  
CC prevented by neutralising the effects of toxin A in the  
CC gastrointestinal tract. Examples are provided of the production  
CC of recombinant C. difficile toxin A in host cells and of the in  
CC vivo neutralisation of toxin A by antibodies against recombinant  
CC toxin A polypeptides. The invention specifically relates to  
CC recombinant proteins derived from Clostridium botulinum toxins  
CC (see AAW68389-400) and their use as immunogens for the production of  
CC vaccines and antitoxins.  
XX Sequence 2710 AA;  
SQ Query Match 53.6%; Score 2783; DB 19; Length 2710;  
Best Local Similarity 52.1%; Pred. No. 1.4e-142;  
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;  
QY 1 MNLVNAQLOKQVYKVFQIEDEYVAIINALREYHNMSSESVVEYKLVKLDNLTNLYL 60  
DB 1 MSLISKEELIKUAY-SIRPRENEYKITLNDYENKLTNNENKYLQKLKUNESIDVFM 59  
QY 61 NTYKESGRNKALKPKFKEYLTMVLELKNLSLTPVEKNLHFIMWIGQINDTAINYNQWKD 120  
DB 60 NKYKTSRRNALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD 119  
QY 121 VNSDITVKEVYDSNAFLINTLKTIVESATNNTLVESFRENLDNDPFDYKNFKRMEIY 180  
DB 120 INAEYNIKLYWDSAEFLVNTLKAIVESSTTEALQLEEEIQNPQPDNKKFYKKRMEFY 179  
QY 181 DKQKHFDYKSOIEENPEFIIDNIITKYLNSNEYSKDLALNKYTEESLNKITANNNDI 240  
DB 180 DRQKRPINYKSIQNKPTVPTIDDIKSHLVSEYRNDRTVLSRYTNSURKNSNHGDI 239  
QY 241 RNLEKPADBDLVRLNQELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQPDLFKS 300  
DB 240 RANSUFTQEELLNYSQELLNAGNLAASDIYVLLALKNFGVGYLDVDMLPGHISDLFKT 299  
QY 301 INKPDSTWTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEDEEVORSFESALSKSKSEIF 360  
DB 300 ISRPSSIGLDRWEMIKLEAIMKYKEYINNTYSNFDKLDQQLKDNFKLIESKSEIF 359  
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDYSQSDVLINQIKNRYKILNDNLNPSINE 420  
DB 360 SKLENLVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVGNRYQFLNQHLPAS 419  
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITYLVKVGFPADVRSTINLSGPGVYTGAYOD 480  
DB 420 DNNFTDTTKIFHDSLFNSATAENSMTLTIAPYLVQGFMPPEARSTISLSPGAYASAYD 479  
QY 481 LLMPKDNSTNIHLEPELNFEPKTKISOLTEOETISLWSFNQARAKSOFEEYKGYE 540  
DB 480 FINLOENTTEKTLKASDLTEFPENNLSQLTEQEINLSWSPDQASAKYQFEKYVRDYG 539



XX OS Plasmodium vivax.  
 XX PN WO9011772-A.  
 XX PD 18-OCT-1990.  
 XX PF 03-APR-1990; 90MO-US01849.  
 XX PR 06-APR-1989; 89US-0334270.  
 XX PR 05-APR-1989; 89US-0334041.  
 XX PA (UYNY ) UNIV NEW YORK STATE.  
 XX PI Barnwell JW, Galinski MR, Wertheimer SP;  
 XX DR N-PSDB; AAQ06328.  
 XX PT Malarial apical end merozoite proteins and peptide(s) - used for  
 PT developing cpds. for treating, preventing and diagnosing malarial  
 PT infection  
 XX PS Claim 2; Fig 1a; 66pp; English.  
 XX CC A MAPP compound having a binding affinity for a Duffy blood group  
 CC antigen of primate red blood cells, is antigenic for the complete  
 CC protein, and may be used in diagnosis, treatment and vaccination  
 CC against invasion by P.vivax and P.knowlesi.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX SQ Sequence 1254 AA;  
 Query Match 5.7%; Score 293.5; DB 11; Length 1254;  
 Best Local Similarity 19.7%; Pred. No. 1.4e-07;  
 Matches 207; Conservative 194; Mismatches 405; Indels 247; Gaps 45;  
 QY 119 KOVNSDYTKVFDVSNFLINTLTKTIVESATNLTLESFRENLDPEFYKPY-----172  
 DB 7 KOTSFDEKKKIEKAVEKMGNTLTK-----ELEKMDDEKNIEKEVEEAQIQYKRIFIDHDVN 62  
 QY 173 -----RKRMEIIVDKQHFIDYKQIEENPEFIIDNIITKYLSEYKSDLEALNKYIEE 227  
 DB 63 LMNDEVEKSLVMEK-----IELYKKEIDELKQ-----KT---NEY-KQDTSNFYTE 107  
 QY 228 SLNKITANNNDIRNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDV 287  
 DB 108 QYNSATQSKA-----KIEQFT-----NIATTK-----KGTSDTSQDI 139  
 QY 288 DILPGIOPDLFKSINKPDSITNTSWEM-----IKLEAMKYKEYIPGYTSKN-----334  
 DB 140 NELESKEEVKRLQLVQKESNWEEMRKQILSMKOLLILNSETAKEISNTQNALGF 199  
 QY 335 -----FDMLEDEQVSFESALSCKSEIFPLPLDDIKVSPLEVKIAFANNVINQA- 386  
 DB 200 RENAKTKLNTKTDLLQORVAAMIEEAKAHKNIDIALEDAQIDTEVSKIEQINREINWKD 259  
 QY 387 -----LISLKD--SYCSDLVINOIKRYKI-----LNDNLNPSI 418  
 DB 260 EIKSYLSEIKYKDKCTEISNKRKGDKIEFLFKPNEESNSKNVINEINENIRNSE 319  
 QY 419 NEGTDF-----NTWKIFSDKLASIN--ENNMMFMKITNYLVKVGEPADVRSTINLS 469  
 DB 320 QYKLDIEDAKQASTKVELPHKHETTISNIFKRESEILGVETKSKQKINKRAEDIMKEIERH 379  
 QY 470 GPGVYT--GAYQDLMFKONSTNIHLEPELRNFEFPKTKISQLTQEITSLMSFNQARA 527  
 DB 380 NSIEIQTVKGFQENLKNLNEPHNYDNAEDLNKDKSTNAKV--LIETNLESV-----KHN 432  
 QY 528 KSQFEYKGYFGALGE-----DNDLPAQNTVLDKDYVSKKILSSMKTRNKEYIHIVQL 584  
 DB 433 LSEITNIKQ-----GEKIYSKAKOIMQKATSENTAEKTLKVKDDQSQSNVNYL----483

QY 585 QGDKISYEASCNLPKDPYSSI-LYQKNIEGS--ETAYVYVADAE-IKEIDKYRIPYQI 640  
 DB 484 --NQITERNLIVTEKRLNGIDSTITNIEGALKESKGNYEIGFLEKIEEIGKRNKRLKVD 541  
 QY 641 SNKRNIKLTFIGHGKSEFNT-----DTFANLD-----VDSLSSSETITLNIAKADI---686  
 DB 542 ITKKSINST-VGNFSSLFNNFDLQYDFNKNINDYENKMGIEYNEFEGLSKISENLRNA 600  
 QY 687 -----SPKYI-----EINLIG-----CMFYSYSIYAETYPGKLLKIKDRVSELM 727  
 DB 601 SENTSDYNSAKTLRLLEAQKEKYNLLNKEEANKYLRDVKVBSF--RFIFNMKESLDKIN 658  
 QY 728 PSISODSITVSANQVEVRINEBGRKREILDHSGKWINKESIIKDISSEYISFNPKENKI 787  
 DB 659 EMIKKEQLTVNCHGVNQLVENIKELVDE-----NLSLDILKQATGKI-----EEIQK 707  
 QY 788 IVKSKYLHELSTLLQEIERNNANSSDIDLEKKVMLTE----CEINVASNIDROIIVEGRIB 843  
 DB 708 ITHSTLKNKAKTILGHVDTSAKYVGIKITPELALTTELLGDAKLKTAQELKFESKNNVLE 767  
 QY 844 AKNLTSDSINYIKNEFKLIESISD-----SLYDLKHQGLDDSHFISFEDISK-TENG----895  
 DB 768 TENMSKNT-----NELDVHKNIQDAYKVALEILAHSEIDITKQ-----KSSKLIEMGNOI 818  
 QY 896 -PRIRPINKETGNSIFITEKEIFSEYATHISK--EISNI--KDTIPDNVANGKLVKYN 949  
 DB 819 YLKVVLINQYKKNKISSIKSEAEVSKTGNVSKHSELSKITCSDKSYDNI-----IA 871  
 QY 950 LDAHEVNTLNSAFFIQ-----SLIEYNTKESLGNLSVAMK-----986  
 DB 872 LEKQTELQNLNRNFTQKTNNTNSDSKLEKIKTDSESLKNAKLTLEGEVNAKASSDNHEH 931  
 QY 987 VQVYAQLPSTGLNTITDASKVVELVSTALDETI 1019  
 DB 932 VQSKSEPNPALSEIEKEETDIDSNTALDELL 964  
 RESULT 7  
 AAW24575  
 ID AAW24575 standard; Protein; 1254 AA.  
 XX AC AAW24575;  
 XX DT 25-MAR-2003 (updated)  
 DT 10-NOV-1997 (first entry)  
 XX DE Merozoite apical-end protein clone 5.3.  
 XX KW Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;  
 KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;  
 KW Duffy blood group antigen; red blood cell; therapy.  
 OS Plasmodium vivax.  
 XX US5646247-A.  
 XX PD 08-JUL-1997.  
 XX PF 04-OCT-1991; 91US-0792865.  
 XX PR 04-OCT-1991; 91US-0792865.  
 PR 05-APR-1989; 89US-0334041.  
 PR 06-APR-1989; 89US-0334270.  
 PR 03-APR-1990; 90MO-US01849.  
 PR 02-NOV-1990; 90US-0608639.  
 XX (UYNY ) UNIV NEW YORK STATE.  
 XX PI Barnwell JW, Galinski MR;  
 XX DR WPI; 1997-362995/33.  
 DR N-PSDB; AAT80072.  
 XX





CC antibody raised to immunogens comprising the sequences of (I), are  
 CC useful in the detection of infection with *P. falciparum*. Furthermore,  
 CC (I) especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent *P. falciparum*  
 CC infection, or they can be used to identify drug resistance in  
 CC *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the  
 CC subsequent identification of proteins encoded by it will help to expand  
 CC our understanding of parasite biology, a process hampered by the  
 CC complexity of the parasitic lifecycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many  
 CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.

XX Sequence 1516 AA;

Query Match 5.4%; Score 280; DB 21; Length 1516;  
 Best Local Similarity 19.2%; Pred. No. 9.6e-07;  
 Matches 230; Conservative 187; Mismatches 351; Indels 432; Gaps 57;

QY 6 KAQLQWVVKRIQOEYVAIILALEEYHNKSESSVVEKYLKLDINLTP-----NYL 60  
 DB 131 KTAELKLVNYQO-----RTLNSMKSKKNKNDNS-----NNIEDKTNPNT 173  
 QY 61 NTYKSG-----RNKALKFKFYL-----TWEVLE 85  
 DB 174 NTQKSNQNTKNTPNKINADISKSLIQIYDDIKEKSLNSLVEHGVNVPVSKDVL 233  
 QY 86 LKNSLTPVEKNLHFI-----WIGQINDT-AINYINQWQVNSDYTVKVFYDSN-AF 136  
 DB 234 ICNDLSKIKNKFIMTDFGPVLFGEQDGMGTVENINKLDNRNKD-----ENNL 286  
 QY 137 LINTLKTIVSATNTLSFRENLDNPFYKFKRMEIYDKQHFIDYKQIE 196  
 DB 287 SINYNKVQVNNNDKDKENINEVRD-----QKNYV-YKNK----- 326  
 QY 197 NPEFIDNIKTVLSNEYSKLEALNKYIEESLKITANGNDIRNLEKFADELVLRL-- 254  
 DB 327 -----ENINNYLDDDEK-----EDQNKGVYNNDDI-----DEQIRK 364  
 QY 255 MARKKYIESIPKTFGLCWRPVDIIDISNTYEMLE-----ISETLKVENKFKQHL 418  
 DB 365 MARKKYIESIPKTFGLCWRPVDIIDISNTYEMLE-----ISETLKVENKFKQHL 418  
 QY 282 GYVLVDVILPGQDPLPKINPKDSITNTSWEMIKLEAMKYKEYIPGYSKDFMLDEE 341  
 DB 419 NVLDENNSTFVNMNLLKNIN-----YK-----KNDLLIEGG 450  
 QY 342 VORSF-----ESALSSKSKDSE-----IFPLDDIKVSPLEVKIAPANNV 384  
 DB 451 EKKSFLINLVDSVSSNSRLNDENERGINMFTNDEKSIINN-----YNNNNNN 506  
 QY 385 QALISLSDVSYDLVINOIKRYKILNDNLNPSINEGTDFTMTKIFSKL--ASISND 442  
 DB 507 NNNNDNNNDNDVIEHNKNNNIYDNKYN-----VECSSEKINDNGISKN 554  
 QY 443 NMFPMIKITNYLVKVPAPDVRSTINLSGPGVTGAYQD-----LMFKNSNTNHLLEPE 497  
 DB 555 --INILEPNNLD-----SNIFLEGKDEYKVYVYVKEIRIPLFKEINKEI----- 599  
 QY 498 LRNFEPKTKISQLTQEITSLW-----SFOARAKSQFEEYKGY--- 538  
 DB 600 ----FEKLPLKQVQIILQDIKEWYTNRIKAKSKDDMDVFSQVQLETYVRMIKTD 656  
 QY 539 -----FEGALGDDDLDFQANT-----VLDKDYVSKLSSMKTRNKEYIHY 582  
 DB 657 KLKIKWAENIQSVEGEL--LINKOLSKNTDINIKDYNVLQK---KSKKKKKFLND 711  
 QY 583 QLQGGKISYEACNLPFSKDPYSILYQKNIEGSETAYYYYVADAETIKEDKVRIPQISN 642  
 DB 712 N-----TYNFTTESKYQDL-----YVKGESKEDIKNQIDPVTQE 746

QY 643 -KRNKILFTIGHKSEFNTDTFANLDVDSLSSEIETILAKADISPKYIEINLLGCNMF 701  
 DB 747 CYRNDIIRDTHDKS-----DIFKNIKDNKKYIYNLEOEINEKK-----NYN 794  
 QY 702 SYSVIAETYPKOLLKLDKRV-SELMPISIQ---DSITVSANQYEVRI-----NEEGKRE 753  
 DB 795 KNNDSNKTFF-----FLKIENEFKOLLDDSDQIFGDSLADIKEYNYTADNLDNNENKS 850  
 QY 754 ILDHSGKWKNEESIIKDISKE---YISFNPENK--IIVKSKYLHELSTLLQEIENNA 808  
 DB 851 LYEDGENFTIRNEPTITNEYEKNIIYISDEQYNEEDIFFDK-----IKEKKNND 903  
 QY 809 NSSDIDLEKKVMTLCEINVASNIDRQIVEGRIBEAQNLTSD-----SINYIKNEFKLIES 864  
 DB 904 TSSD-DFB-----NCSVQ-----EKIYVNEKIEYNNKNDKSSSSSIILEIKYKKE 951  
 QY 865 ISDSLY-----DLKHQGLDSDHP----- 883  
 DB 952 KKDELVSPNLCVLLDDEFHSNDLENNYISVSSDDMKTNVSKNITGVKENKVDKTNVEYD 1011  
 QY 884 -----ISPEDISKTEGPRIRFINKETGNSIF---IETEKEIFSEVATHISK--- 927  
 DB 1012 KKGDDGVTEISFEDSHKLEES---KF---DNNNIYDNDDELEKNLSKDYISDVKQHN 1065  
 QY 928 EISNIKDTIFDNVNGKLVKVNLDAAHEVNTLNSAFFI-----QSLIETVNTTSESLSNLSV 983  
 DB 1066 NIYNIERGEDBRENEFVENKIQSTESHKSNE-----FICTENKSLRKQYMSKEDISNVR 1120  
 RESULT 10  
 AAB18294  
 ID AAB18294 standard; Protein; 980 AA.  
 XX AC AAB18294;  
 XX DT 07-NOV-2000 (first entry)  
 XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.  
 XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
 XX OS Plasmodium falciparum.  
 XX PN WO200025728-A2.  
 XX PD 11-MAY-2000.  
 XX PF 05-NOV-1999; 99WO-US26796.  
 XX PR 05-NOV-1998; 98US-0107131.  
 XX PA (HOFF/) HOFFMAN S.  
 XX PA (CARU/) CARUCCI D.  
 XX PA (GARD/) GARDNER M.  
 XX PA (VENT/) VENTER J C.  
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;  
 XX WPI; 2000-365347/31.  
 XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
 XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 XX PT diagnosis of *P. falciparum* infection -  
 XX PS Disclosure; Page 362-365; 577pp; English.  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*.  
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
 CC vaccines against *P. falciparum* infection comprising (i) or (ii).  
 CC (i) and (ii) are useful for the development of vaccines against

P. falciparum infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with P. falciparum. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent P. falciparum infection, or they can be used to identify drug resistance in P. falciparum. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

XX Sequence 980 AA;

Query Match 5.2%; Score 272; DB 21; Length 980;  
Best Local Similarity 20.3%; Pred. No. 1.5e-06;  
Matches 226; Conservative 180; Mismatches 393; Indels 316; Gaps 52;

QY 6 KAQLQRMVY-VKPRIOEDYVAILNALAEYHNHSSSVVEKYLKLDINNLTNYLNTYK 64  
DB 39 KEKGEKVYELKEKLONDE--KLINDLKKNSY-----QYVKMDYKRENNLINEIN 89

QY 65 KSGRNKALKFKFYLTWEV---LELKNNSLTPV-ERNLHFPIWIGGOINTAINYNQWMD 120  
DB 90 K-----LKLFIENKMTVERGNEMNNKLEEMKQKNKELINNLNLDISDELKNCIRQVNS 143

QY 121 VNSDYTVKVVYDSNAFLN-----TLKKT1-----VES-----ATNNTLESFR 158  
DB 144 VSRN--MANVEKEKEWIIINLOILRMKNDTMRKISKFVEQEKVLPFKLYTLNNDIFSKN 201

QY 159 ENLNDPEFDYKPRKREMEIYDQKHFD-----YVKQSIEENPEFIIDNIITK--- 208  
DB 202 EKLDN-----MQKLDNDVNEKYNIVECLNNYKTEHKEQIEKKIERI--NTLQNYV 251

QY 209 YLSNEYSKOLEANKYIE-----ESLKNITANGNDIRNLEKFAEDLV 252  
DB 252 YLKKEYDLANKLEKNEIHEGKLEHLSHCYEENQKLENEIKRNSFIKNKDR--KIDLL 309

QY 253 RLYNQELVERWNLAAASDILRI SMLXEDGGVLDVILPGIQDLPKSIKPKDSINTSW 312  
DB 310 TNENELKK-----KEINNIKMEKQNVIRKNEQLLDKIDKENERKMEHVNLQN--- 360

QY 313 EMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSDKSEIFLPDDIKVSPLE 372  
DB 361 ELIKRE-----LQNKCSKDIEFCKEKEDK-----IKNLEDDLE 396

QY 373 VKJAFANNSVINQALISLSDYSCLVINOIKNRYKILNDNLNPSINEGDFNTWKIFS 432  
DB 397 KKKCIEN---LKDDELINIK-----KQMEDKMHMTNE-MDLLSNKVEELNKRINTYE 443

QY 433 DKLASTSNE--DNMFMIKITNYLKVGFAPDVRSTINLSGPGVY--TGAYQ-----DLL 482  
DB 444 KNIVELNDELVDVTKKLNDEFLK---EEKKKNIDM-----VYKIKYEYIQLKEKENEID 496

QY 483 MFKDNSTNIHLLPELNFPPKTKISQLTEQITSLWSPNQARAKSQFEYKKGFEGA 542  
DB 497 SLKQNEQNLHLVXNE-----ELNEKEIILKNKYDK-EINMIEQYNKKI----- 539

QY 543 LGBDDNLDFAQNTVLDKDYVSKILSSMKTRNKEYIHYI-----VOLQGDKISYEASCNL 597  
DB 540 ---QEEKDMLNNIKSMQDTHQNIEMQENKELKRLKNCVCDMLQSLIKENKEM 596

QY 598 FSK-DPVSSSTLYOKNIEGSETAYVYVADAEIKS-IDKYRIPYQISNK----- 643  
DB 597 QEKVEEYKNLKKQK-----DOELKNIIQEYDERIEIQNKEMEDIVNCEEK 642

QY 644 -----RNKILPTIGHGKSEFNTDTFANLDVDSLSSEIETILNLAKADISPKYIEINLLG 697

DB 643 LKQAKINNKKLT-----TATNMANNMMLMDENLKEKDKKINDLWKOMEKKKEINKLV 696

QY 698 -----CNMFYSIVABETYP---GKLLLKIKDRVSELMWPSISQDSITVSANQ 741

DB 697 EEKSKLEHSHVKIQNEMSLLVQNEKLEEMGLSRIAKD-MEEIKKOMEKYEEEKKKNVE 755

QY 742 YEVRIINEGKEIRILDHSGWINKESIIKIDISSKEYISFNPKNENKIIIVKSKYLHELSTLL 801

DB 756 EERKNEERKKNVEERKK--NEEEKKKNVEERKK--NEEEKKLEKDKHQFE----- 804

QY 802 QEIRNANSSDIDLEKKVWLTECEINVASNIDRQIVEGRIEAKNLTSDSINYKNKFKL 861

DB 805 -----BEKERM---EYEHQKEDR---KRDKKKKKGHSSDKEEKYNKKEKT 844

QY 862 IBSISDSLYDLKHQGLDSDSHFISPEDISKTEGFRIFRPFINKETGNSIFIETKEIFSEY 921

DB 845 KEKSSNLIIF-----DEEYIIQEEL-----RDTGNCFI-----Y 874

QY 922 ATHISKEISNIKDTIFDNNVNGKLVKKNLDAHAHEVNTL-----N 960

DB 875 LKSLSEKL-----DVINKLKSDDALLN-DAFNKINLAITSWNIPNEENKEGDNITTVEN 929

QY 961 SAFFIQSLIEYNTTKESLNSLV--AMKVQVYVYAO 993

DB 930 TATEGNITIDTTEVMNNEEVYKIFSVEKYDML 964

RESULT 11  
AAB18241  
ID AAB18241 standard; Protein; 1121 AA.  
XX AAB18241;  
XX 07-NOV-2000 (first entry)  
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:98.  
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoasidae; infection; insecticide.  
XX Plasmodium falciparum.  
OS WO2000025728-A2.  
PN 11-MAY-2000.  
PD 05-NOV-1999; 99WO-US26796.  
PF 05-NOV-1998; 98US-0107131.  
PR (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
PS Disclosure; Page 229-232; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,



(I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAH70078 to AAH70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

Sequence 1121 AA;

Query Match	5.1%	Score 266.5;	DB 21;
Best Local Similarity	20.2%;	Pred. No. 3.5e-06;	
Matches 238;	Conservative 183;	Mismatches 372;	Indels 385;
			Gaps 55

Qy	5	NKAOLOQXVVYKFRIOE--DEYVAIINLAEEYHNSE-----SSVEKYLYKLKDI	52
Db	41	NEKRTNKIKYKSAQSLFDKGLNIHDKLILFKNLPKYKCAKYECSIAKEVYKVL----	95
Qy	53	NNLTDNYLNTYKSGRNKAL-----KKPEY-LTMEVLELKNSLTPVEK-	96
Db	96	-----LDEYKXCFNYISLCOIIQSVKIFDELDKTFTDYNYFIEYGVNDKRVLNKINBEI	148
Qy	97	-----NLHFITWGGQINDTAINYINO-----W-KDVN-----SDYTVK	128
Db	149	YFKWCDTFFHRREILGKCNKIMSIYHGGWNGELHFLIYFRWNQDNKLLIFYNYIFN	208
Qy	129	FVYDSNFLINTLAKTIV--ESATNNTLESFRENLDPEFDYKPYKRMIEIYDKQKH	185
Db	209	YVFDHMYLFNHEIYKLFIPFNKYLNNNSNIPFNKLIQ-EMEFNLYYFREIK-----NEKY	263
Qy	186	PI-----DYK--SQTEENPEPTIDNIIKTYLSNEYSKOLEALNKYIEESLNKITANN	236
Db	264	YIIKWNKKEIYKCKFAKPHENVDH-IDN-----EKILNILURLYVDNSILDOIDINN	312
Qy	237	GNDIRNLEKFADEDLVRIYQBELVERWNLAAASDILRI--SMLXEDGGVYLDVDILPGIO	294
Db	313	-----XMLCNLNNNLN--NIEVSKLLNFYCTLIKK--GKY-----	346
Qy	295	PDLFKSINKPDSIINTSWEMIKLEAIMKYKEVPCYVTSKNFMDLDEEVQSPESALS	354
Db	347	-----DNDWTIYKLEKVIKATHILCDKTKNLETFCSDI--DYSTLLNSLN	390
Qy	355	DKSEIFLPLDDIKVSPLE-----VKIAFANNSVINOALISLKDSCYSDLVINOIKRYK	408
Db	391	NKPIINKIIDKNFILFYECCLKILNIFKVPFQSLCISLISLKNY-----YNIURNVY	445
Qy	409	ILNDLNPSINEGTDNFNTMK-----IFSDKLASISNEDNMFMK--ITNLYLKVGFA	459
Db	446	IVNVVL-----FNDIMKFSLYLCNIFLGKRIKTENENAVLIHNNDOQNYSN----	492
Qy	460	PDVRSNTINSLGPGVYTGAYQDLLMPKDNSTNIHLEPELRNPEFPKTKISQLTQEITSL	519
Db	493	-----KENIKDI-IIQKRIKEYIFYK-----	512
Qy	520	WSPNOARAKSQPEEYKGYFEGALGEDDNLDPAQNTVLDKDVSKILSSMKTR-NKEVI	578
Db	513	-----MENYKDPHFK-----LKSDSLLSILKLSNTFVKINVEVYN	546
Qy	579	HYIVLOQDKISYEASCNLFS-----KDPVSSIL-----YQKNIEGETAY	619
Db	547	SYDFYLLFNNI-----SCLINFLVNRNSVKYKOTYIYLNDLSFYVKIKNDRTKKKK	602
Qy	620	YYYVADAEIKBIDKYRIPYOISNKNKIKLFTFGHGSBFNTD-----TFAN-----LD	667
Db	603	NFFLLSSSMKELICWNI-LSVNR-----YKHLHEEDNFDQKQOYVCSLFTLNNLPD	655
Qy	668	-----VDSLSSEITELINLAKAD--ISPKYIEINILGCNMNFSYSI--YABETPYGKLL	716



XX The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.

XX Sequence 1712 AA;

Query Match 5.0%; Score 258.5; DB 21; Length 1712;

Best Local Similarity 20.3%; Pred. No. 1.7e-05;  
Matches 219; Conservative 163; Mismatches 394; Indels 301; Gaps 52;

QY 34 YHNMSESSVYKYLKQI--NNLTDNYLNTYKSGRNKALKKFKYLTMEVLEKNNSL 91  
DB 48 YNN-----VDKIPNKEIYNNIKSNNIQVRVQNNNEKKKEEANYT-----CVNNKY 96  
QY 92 TPVERKULHIWGGQINDTAIYI-----NQWQVNSDYTVFVYDSNAP 136  
DB 97 VTLLKNKVH---VNKYVNSNINKITVPIIKCSNYKIKNNPISHLKSNYENKFKLSN-- 151  
QY 137 LINTLKK--TIVESATNTTLESGFR-ENLNDPE-----FDYKPF---YKRMELIYDKQK 184  
DB 152 -PSNINKGSHKNDVINETMDQHKSEQLANDNIIKLLYDYCIFREDTIKTNTISYNNKN 210  
QY 185 HFIDYKQIOENPEFIIDNIITYLSN--EYSKOLEALNKYIEESLNKITANNNDIRN 242  
DB 211 SFKDN-----EENINTMDNNIKSNSSSYCSYS-----NKNQNNVNHHTL 251  
QY 243 LKFEADELVRVYNQELVERWNLAAASDILRISMLKEDGGVLDVDILPGIQDLPKSN 302  
DB 252 KTEFLNEKNSHTQNEQ-----SIPLLDGLQ-----NNHN 280  
QY 303 KPDSINTTGWEMIKLEAIMKYEYIPGYTSKNFD---MLDEEVORSFESALSSKSDKE 358  
DB 281 SATKFNNIYD--NNSLVNYK-----SOKGIDLHNKMKIETDKNGIITLEKKGHDK 332  
QY 359 IFPLDDIKVPLEVKIAFANNNSVINOALISLKSYSY--CSDLV-----INQIKN--R 406  
DB 333 YY---NNIFLPLNDN---SNNVITTC--DNKESYRNTSDMINKIFEKMWNEKKNILK 384  
QY 407 KYLNDNLNPSINEGTD--FNTTMKIFPSDKLASISNEDNMFMKITYLKVGFAPDVS 464  
DB 385 MKNFNDVIKKKITMAKEKILNSNTTNMKKVSFYNKSDLEDLFNEKENSYS-KYGVKRENGE 443  
QY 465 TI-----NLSGPGVYTGAYODLLMPKDNST--NIHLEPELRNFEPP-----504  
DB 444 DINVIKNKRNININDNDNIIKNDVSKRNIIHNKKKRDDEDFPNNSSAGLLDFFDL 503  
QY 505 -RTKISQLEQEITSL-----WSFNQARAKSQFEYKKGFEKALGED-----546  
DB 504 CRRKVLIELKNVQSSKKKKNKILTNHNSNDQNCHSSDNQCHSSDNQCHS 563  
QY 547 -DNLDPAQNTVDKQVYSKLLSSMTRNKEYIHYIVQLQGDKIYSYASCNLFPSKDPYS 605  
DB 564 SDNQNCDSNACNKKDEKRRKKKKIKKKNK-----KNKSNKSKN--KRETKSK 611

QY 606 ILYQKNIEGSETAYYYYVADAEI--KEIDKVRIPYQISNKRNIKLTPIGHGSKSEFNTDTF 663  
DB 612 KISNNNNNDNMNQDNMGDQRINNENMDQNV--NIQNEGN-----GFNNNNKNNDDL 663  
QY 664 -----ANLVDLSLSSEIETILMLAKADISPKYIEINLLGCNMFYSIYAEETYPGKLLLK 718  
DB 664 NYIYSPNMINHLSSTCE-----KKNKEDNKNDNKPLNS-----SSKMK 703  
QY 719 IKD---RVSELMPISODSITVSANOYEVRIEKGKREILDHSGKWINKESIIKD--I 772  
DB 704 IPEITNNSNEKIVNVSNDEMUVYHNLTVLVNKBQ-----GVTEESSICKRTYF 753  
QY 773 SSKEIYSPNKENIIVKSKYLHELSTLQEIIRNNANSSDIIDLEKKVMLTE-----823  
DB 754 VQFYDSYNNRNEKITDDNMQVEDIYVVKENIKRTLKGDGHDVKTNNMLSEDNSYASGLW 813  
QY 824 -CEINVASN-----IDRQVEGRIEAKNLTSDSIYIKN-----857  
DB 814 GNEINFISNNENCLNSYDISCDERKIPNBEEOBELCSNNI-LVKDIEBKQMGKLPFEE 872  
QY 858 --EFKLIESISDSLYDLKHQNGLDSDH--FISFEDI---SKTENGFRIRFINKETGNS 908  
DB 873 ICVFRINEKNEGHENLRKNNHNDTHKMYSSYEMIQNINKOSTPFCCKDEMEKSGQTN 932  
QY 909 IFIE-----TEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAAEVYN 957  
DB 933 LFYDNYINSVDITKLELNKNCYQHINYEQN-----LIKKENSYAA-ENN 976  
RESULT 14  
ABB54799  
ID ABB54799 standard; Protein; 1090 AA.  
AC ABB54799;  
XX  
XX 16-MAY-2002 (first entry)  
XX Lactococcus lactis protein yoiA.  
XX Lactococcus lactis protein yoiA.  
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX Lactococcus lactis IL1403.  
XX PR2807446-A1.  
XX  
XX 12-OCT-2001.  
XX  
XX 11-APR-2000; 2000FR-0004630.  
XX 11-APR-2000; 2000FR-0004630.  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX WPI; 2002-043418/06.  
XX New nucleotide sequence useful in the identification of Lactococcus  
XX lacticis and related species -  
XX  
XX Claim 6; SEQ ID No 1501; 2504pp; French.  
XX The present invention is related to a Lactococcus lactis nucleotide  
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
XX nucleic acid sequence is useful in the detection and/or amplification of  
XX nucleic acid sequence, particularly to identify Lactococcus lactis or  
XX related species. The proteins of the invention are useful for the  
XX biosynthesis or biodegradation of a composition of interest. The  
XX invention helps research in lactic bacteria, particularly useful in the  
XX production of yogurt and cheese.  
XX Note: The sequence data for this patent is based on equivalent patent  
XX WO200177334 (published 18-OCT-2001) which is available in electronic  
XX format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX	Sequence	1090 AA;	
SQ	Query Match	4.9%; Score 256; DB 23; Length 1090;	
	Best Local Similarity	19.5%; Pred. No. 1.3e-05;	
	Matches	226; Conservative 185; Mismatches 370; Indels 378; Gaps 59;	
QY	11	KMYVVFRIQEDRY-----VAIINALEYHN--MSESSVVEKY-LKLKDINNLTNDVNTY 63	
DB	91	KKAYLKNNLNKDFNLLSKQADKQDYNSEMIKYEEVKKYEQSLUKEYQETKEYENVI 150	
QY	64	KKSGRNKALKKFKYELTMEYLEKNNLSTPVEKNLHFIWIGQINDTAINYNQWKDVS 123	
DB	151	--NDRKAAQKPEE-----NSLALKN-----INDEVQNLIDK-KSVKN 185	
QY	124	DYTVKVFYDSNAPLINTLTKTIVESATNTNLTESFRENLDNPEFY---NKFRKRMEIY 180	
DB	186	D-----AF-IEIKEKLAVEQEEFKAAES---QLNGIVSVYELLKNYQEKKATPE 231	
QY	181	DKQKHFDYVKSQIENPEFIIDNIITKTVLSNE-YSKDLEALNKYTEESLKNITANNND 239	
DB	232	ESRKY--RHANTIKDWEMLDKFEKTSNKEKFEQNLQADDKVKATEEE--ANISNE 287	
QY	240	IRNLEKFADEDLV-----RLYNQELVERNNLAAA---SDILRISMLKEDGGVILD 286	
DB	288	IYN---FVPSLVAMTEKYEHKKVYEAAL-AYNISKKGYDAEYAKFSELKEN----- 337	
QY	287	VDILPGIQDLPKSIKNNKPSITNTSWEMIKLEAIMKYEYIPGYTSKNFMDLDEEVORSF 346	
DB	338	-----FQEIIN---SLKNPSQTVI-----GY-----ITRDM 359	
QY	347	ESALSCKDSEIFLPLDDIKVSPLEVKIAPANN--SVINQALISLSDYSCSDLVINOIK 404	
DB	360	EAKNRVDNLAKDFLEEE-----KFNLOKQIFEQISTSYVLVKN-----LVDEKK 405	
QY	405	NRKYILNDNLNPSINEGTFNMTMKIFSDKLAGISN---EDNMFMIKITNY--LKVGEA 459	
DB	406	GRSKVSGNTLTETLNAQKKTWLVLEVKSKYGLLNHORGSEAFKVIQQVYEEKEFEQ 465	
QY	460	PQVRSINISGPGVYTGAVQDMLFMKDNSTHILPELRNFFPKTKISQLTE----- 513	
DB	466	KEQSFSKIK-----EYTAPETRYNSTK-----ESYDIAKEKLAQSEENLEIVK 510	
QY	514	---QELTSLSMFSNQARAKSFQYKKG--YFEGALGEDDNLDF-AON--TVLDDKV--- 561	
DB	511	KAYSDFIAGLRVATVTKESILEEYKSEKFNLTDTKGEFEVQNNFTGLDETVOKE 570	
QY	562	-VSKKILSSMKTNRKNEYIHYI-----VOLQGDKIS--YEASCNLFSPKPYSSI 606	
DB	571	QLCLNLTEQAQKTHDDYPOYSHSYKTLPEYTKVNSQLEKLSQQYEVWFNDYQNLKSS 630	
QY	607	L-----YQKNEGSETAYYYVADAETKE----- 630	
DB	631	IVRTOYQEQEQ-----RYNADQELYNQVSEBHFCHLDLRNVLKKDNEFETALNNLQ 683	
QY	631	-----IDKYRIPYQISNKRNIKLTFLGHKSEFN--TDTFANLD- 667	
DB	684	FIPKNTLTPPIGEATGSNFDIFENPTEIVSESSKLT-----NKFELQWTTTYQNLVA 739	
QY	668	-VDSLASEI-----ETILNLAKADISPKYIEINLGLGNMFSYSIYA 707	
DB	740	LVDSISKELQALTEYQNLQSIFIKVGGKRVLDGNSDIAYR-----GTNI----- 785	
QY	708	ETYPKGLLKIKDRVSELMPSISQDSITV-SANQY-----EVRINE-----EKGREI 754	
DB	786	-ESYLERLAINFKWEQSEYVKKRQVFSLLVKNYRQYLEESVKLNENIARMSKNGLQEI 844	
QY	755	LD-----HSGKWINKK-----ESIIKDISKVEISFNPKENIIVKSKYHLSTL 800	
DB	845	INAINSLPNYMDIYSTSYRVELLALGALGAEVSDLTLHR-NIDELKEINKVTFEHLAL 903	
QY	801	L-----QEIRNNANSSDIDLEKKVMLTECEINVASNIDRQIVGEIRB---EAKN 846	

DB	904	IDENLNTINHPSEIRSSVDQSIINLRKQ--LDEMEKITGIPTRKVMIESSLPNDLDLKS 961	
QY	847	LTSDSIN-YIKNEFKLIESISDSLDYLDKHQGLDSDHFTSPEDISKTENGFRIRFINKET 905	
DB	962	YFTGALNENVKOLFSTIISSSSFINSVKKVN-----QFKRVQNLLSV---NEG 1006	
QY	906	NSIETIETEKEIFSSYATHISKEISNIKDTIPDNVNGKLVKKVNDLDAAEVNTLSAF-- 963	
DB	1007	GNAIILPEVTPADPKKELEFSLVIATIK-----NIDAPVKVEKLESAYVK 1051	
QY	964	-----FIQSLIE 970	
DB	1052	PEKPENLADKPSYVEDIVE 1070	
RESULT 15			
ID	ABP73774		
XX	ABP73774 standard; Protein; 1948 AA.		
AC	ABP73774;		
XX	30-JAN-2003 (first entry)		
DE	Candida albicans essential protein SEQ ID NO 7611.		
KW	Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;		
KW	signal transduction; DNA replication; cell division; growth;		
KW	proliferation; Candida albicans; fungicide; antifungal.		
OS	Candida albicans.		
XX	WO200253728-A2.		
XX	11-JUL-2002.		
XX	26-DEC-2001; 2001WO-US49486.		
XX	29-DEC-2000; 2000US-259128P.		
PR	20-FEB-2001; 2001US-0792024.		
PR	22-AUG-2001; 2001US-314050P.		
XX	(ELIT-) ELITRA PHARM INC.		
XX	Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;		
XX	WPI; 2002-566694/60.		
DR	N-PSDB; AB232324.		
XX	Constructing strains for identifying gene products as effective targets		
PT	for therapeutic intervention, by inactivating in the strain one allele		
PT	of a gene and placing other allele of the gene under conditional		
PT	expression		
XX	Claim 44; SEQ ID NO 7611; 167pp + Sequence Listing; English.		
XX	The invention relates to constructing (M1) a strain of diploid fungal		
CC	cells in which both alleles of a gene are modified, comprising modifying		
CC	one allele by insertion or replacement by a cassette having an		
CC	expressible selectable marker and modifying other allele by		
CC	recombination, of a promoter replacement fragment with a heterologous		
CC	promoter, so that expression of the second allele is regulated by the		
CC	promoter. (M1) is useful for constructing a strain of diploid fungal		
CC	cells in which both alleles of a gene are modified. The diploid fungal		
CC	cells having both alleles modified are useful for identifying a gene that		
CC	is essential to the survival or growth of a fungus, a gene that		
CC	contributes to the virulence and/or pathogenicity of a fungus, a gene		
CC	that contributes to the resistance of a diploid fungus to an antifungal		
CC	agent, an antifungal agent that inhibits the growth of a diploid fungus		
CC	and for identifying a therapeutic agent for treatment of a mammalian		
CC	disease. (M1) is useful for identifying a compound which modulates the		
CC	activity of a gene product, preferably enzymatic activity, carbon		
CC	compound catabolism, biosynthetic, transporter, transcriptional,		
CC	translational, signal transduction, DNA replication and cell division		

CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: the sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.  
XX  
SQ Sequence 1948 AA;

Query Match 4.9%; Score 256; DB 23; Length 1948;  
Best Local Similarity 19.3%; Pred. No. 2.7e-05;  
Matches 233; Conservative 207; Mismatches 441; Indels 324; Gaps 53;  
QY 4 VNKAOLQKVVYKPRIQEDYVAILNALBEYHNMSESSVVEKYKLKQINNTDYNLNTY 63  
DB 273 INYULEKSRVISQSNQERNYHIFQFLKGYENLPSLGL-----NKOMS--TYNLSGG 324  
QY 64 KKSGRNKALKKPEYLTM-----EVLKKNLSLTPVEKNLHFIWGGQI----- 107  
DB 325 KLITNN--VDDPKFENLLVEAFKINGFTTNKNFIYQVLAILHLGNITFTSMKSEQANF 382  
QY 108 -NDTAINYNQKQVNSDY-----TVKFVYDSNA----- 135  
DB 383 TNDSPIDRICELLAVDKOLFQVQLLRPKVKAGREFITKSKPNEVKFAIDAPAKLYEKL 442  
QY 136 --FLINTLKKTIVES-----ATNLTLESFRENLD---PEFDYKPYRKMEIIVDK 182  
DB 443 FQFIITKINEKUDQHGNDNTANNNDNNFFIGVLDIAGFEIFDINSPEQLCINYNTEK 502  
QY 183 QRHFIDYKYSQIEENPEFIIDNIITYLSNEYSKOLE-----ALNKYIEES-- 228  
DB 503 LQOFFNHH-SFIEQSEYLRNINNEFI--DFGQDLQPTIDLIETKQPMGILKLDEECL 559  
QY 229 -----LNKITANNNDIRNL--EKPADELDVRLYNQEL--VERWNLAASDILRI 274  
DB 560 MPKSDASPMKSLKNFTTHKFKSENKPKNGFIHHYAGKVEYNVENW-LQKNTDPISE 618  
QY 275 SMLKEDGGVYLDVILPGIQPLPKS-----INKPDSITWTSWEMIKLEAIMKYKEYI 327  
DB 619 SIL-----NLLPDSQNELIADMFINDPHINRPO--TNGGNSKLKTASQKHQOL 666  
QY 328 PGYTSKNFMDLDEEVQSPESALSCKSEIFPLDLDIKVSPLEVKIAPANNVINQAL 387  
DB 667 KTLMDQ--LESTEPHFVRCILPNLEKANKFDKNLVGLQRC-----NGVLEGIR 714  
QY 388 ISLKDSYCSDLVINQIKRYKILNDNLNPSINEGTFDNTMTKIFSDKLASINEONMMFM 447  
DB 715 IT-RAGYPNRMFDEBFQIYYSIICNELSSPQNKTNCEIIL-----KFVKLPNEDFKVGL 768  
QY 448 IKITVYLVGPAPDVRSTINLSGPGVYTCAYQDLLMFKDNSTNIHLLEPELRFEPPTK 507  
DB 769 TKI--FFKNGILGKLEIIRDLAKNIFTDLQK--VIRGNLRLVL-----KOK 812  
QY 508 ISQTEQEITSLWSFNQARAKSQFBEYKGYFEGALGED-----DNLDPAQN-----TVLD 558  
DB 813 IKEIQSAQIISRTWYLDLEIKNSPWRMLFFHVKPLEDSAKVLSKQLENLQTLTVKL 872  
QY 559 KDYVSKKILSSMKTRNKEYIHVIVQLQDKIISYEASCNLFSDKDPYSSILYQKNIEGSETA 618  
DB 873 KD--SEKLTGKLETONKLEQNMNLDQEMINITSIAK--EKD--DKLSQLRLTENSK 925  
QY 619 YYYVADAEIKEDKY-----RIPYQISNKRNIKLTFFIGHGKSENT-----DT 662  
DB 926 HRIETLEIKLDPKQFQNDLINEHEKLTTRSLLELHDKHNAKV-----EELNSLNKLHD 978  
QY 663 FANLDVDSLSSBIETILNLAKA-----DISPKYIEINLLGCNM----- 700  
DB 979 SAQSELSNLKQQLKEITKLNQAHSKELLKFKQWHDKSIEDLNEKVSRLNKLKQADLN 1038  
QY 701 FSYSIYAETYPGKL---LLKIKDRVSELMPSISQDSITVSANQVVRINEEGKREILDH 757  
DB 1039 SSKDVVISSE--HGKLQNEIMKLKELCSDYERKQKQORTIDSLQIQLKDEDYK----- 1090

QY 758 SGKWKNEESIIKDSSKEYISFNPKENKIIIVKSKYLHELSTLLQEI-RNNANSSDIDLE 816  
DB 1091 -----SRVSEKIEEAREKVTLKKGVEKKSQEIQYKSEIKKLAELASSNTKITQLDOS 1145  
QY 817 KKVMLTECEINVASNIDROIVEGRIEERAKNLTSDSINYIKNEFKLIES-----ISDSLVDL 872  
DB 1146 TKEL-----SSLKSNESKYVAE--IESAKQQLSKKI-----BEYSSIEADYNRLQTELKQM 1194  
QY 873 KHONCLDSDHSPISPEDISKTEGPRF-----PFINKETGNSIFETEKE-----IFSEVA 922  
DB 1195 KQTN-----TBYSTRITELSTKLQNVTEAKSKEIEKENQPNPQFMEEF 1239  
QY 923 TH-----ISKEISNIKDTIFDN-----VNGKLV-----KKVNL-----DAAHE 955  
DB 1240 THMKLVNVEQTASLRKEKEPENKKLSEELQMLKERIMNGSLTSMDLTPKRSLSAIGDKSMI 1299  
QY 956 VNTLSAFAFFIOSLIEYNTTKESL-----SNLSVAMKVQVYVAQFSTGLNTI 1001  
DB 1300 TNTVDS-----FNKEIENLKFQLOEQGNFORAENYAIELQKLNKLLTTRGLNTN 1350  
QY 1002 TDASK 1006  
DB 1351 TDYEK 1355

Search completed: November 5, 2003, 19:29:28  
Job time : 36.75 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:27:53 ; Search time 14.25 Seconds  
(without alignments)  
3028.567 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1020  
Perfect score: 5192  
Sequence: 1 MNLVNAQLQKVMYVKFR1Q.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents RA:\*

- 1: /cgn2\_6/prodata/1/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/prodata/1/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/prodata/1/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/prodata/1/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/prodata/1/iaa/6CTUS COMB.pep.\*
- 6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4086	78.7	2366	1 US-08-480-604A-10	Sequence 10, Appl
2	4086	78.7	2366	2 US-08-405-496A-10	Sequence 10, Appl
3	4086	78.7	2366	3 US-08-915-136-10	Sequence 10, Appl
4	4086	78.7	2366	4 US-08-957-310-10	Sequence 10, Appl
5	4086	78.7	2366	4 US-10-011-366-10	Sequence 10, Appl
6	2783	53.6	2710	1 US-08-480-604A-6	Sequence 6, Appl
7	2783	53.6	2710	2 US-08-405-496A-6	Sequence 6, Appl
8	2783	53.6	2710	3 US-08-915-136-6	Sequence 6, Appl
9	2783	53.6	2710	4 US-08-957-310-6	Sequence 6, Appl
10	2783	53.6	2710	4 US-10-011-366-6	Sequence 6, Appl
11	411.5	7.9	3169	4 US-09-453-702B-257	Sequence 257, App
12	261.5	5.0	2184	4 US-09-417-485B-6	Sequence 6, Appl
13	247.5	4.8	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
14	237.5	4.6	3433	4 US-09-091-501B-10	Sequence 10, Appl
15	235	4.5	3248	1 US-08-353-700-1	Sequence 1, Appl
16	235	4.5	3248	5 PCT-US95-16216-1	Sequence 1, Appl
17	224.5	4.3	1151	4 US-09-134-001C-3242	Sequence 3242, Ap
18	220	4.2	3696	4 US-09-134-001C-5080	Sequence 5080, Ap
19	218	4.2	2285	4 US-09-308-375-2	Sequence 2, Appl
20	215	4.1	2662	4 US-09-595-684B-31	Sequence 31, Appl
21	211	4.1	956	4 US-09-134-001C-4452	Sequence 4452, Ap
22	209.5	4.0	2504	4 US-09-328-352-5821	Sequence 5821, Ap
23	209	4.0	1093	3 US-09-315-793-52	Sequence 52, Appl
24	208.5	4.0	1786	3 US-08-973-462-8	Sequence 8, Appl
25	206.5	4.0	1010	4 US-09-134-001C-5178	Sequence 5178, Ap
26	204	3.9	3878	4 US-09-914-259-11	Sequence 11, Appl
27	202.5	3.9	1169	4 US-09-255-829-20	Sequence 20, Appl

28	202	3.9	840	3 US-08-974-549A-190	Sequence 190, App
29	202	3.9	872	3 US-08-851-843A-8	Sequence 8, Appl
30	202	3.9	872	3 US-08-851-843A-54	Sequence 54, Appl
31	202	3.9	872	3 US-08-974-549A-221	Sequence 221, App
32	202	3.9	872	3 US-08-854-050-8	Sequence 8, Appl
33	202	3.9	872	3 US-08-854-050-54	Sequence 54, Appl
34	202	3.9	872	4 US-09-430-323-8	Sequence 8, Appl
35	202	3.9	872	4 US-09-430-323-54	Sequence 54, Appl
36	200.5	3.9	859	1 US-08-053-614-2	Sequence 2, Appl
37	200.5	3.9	859	1 US-08-316-397B-2	Sequence 2, Appl
38	200.5	3.9	859	2 US-09-034-306-2	Sequence 2, Appl
39	200.5	3.9	859	3 PCT-US93-09782-2	Sequence 2, Appl
40	200.5	3.9	859	5 PCT-US93-09782-2	Sequence 2, Appl
41	200.5	3.9	1181	1 US-08-053-614-4	Sequence 4, Appl
42	200.5	3.9	1181	1 US-08-316-397B-4	Sequence 4, Appl
43	200.5	3.9	1181	2 US-09-034-306-4	Sequence 4, Appl
44	200.5	3.9	1181	3 US-09-259-437-4	Sequence 4, Appl
45	200.5	3.9	1181	5 PCT-US93-09782-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-480-604A-10  
; Sequence 10, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027







```
QY 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCYSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSGDMEASPLEVKIAFANNVINOALISLKDSCYSDLVINOIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKITYLVKGFAPDVRSTINLSPGQVYTCAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMELGKLVGVFPDVKTTINLSGPEAYAAQD 480
QY 481 LLMFKDGNSTNIHLPELNPFPKTKISQLTQEITSLWSFNQARAKSQFEYKGYE 540
Db 481 LLMFKEGSMNIHLIEADLNFEISKTNISQSTEQEASLWSFDDARAKAQFEYKRYE 540
QY 541 GAGGEDNLDFAQNTVLDKDYVSKKILSSMKTRKEYIHYIVOLQDKISYEASCNLPSK 600
Db 541 GSGLEDNLDPSQNVVDKEYLLEKISSLARSSERYIHYIVOLQDKISYEACNLFAP 600
QY 601 DPYSSILYQNVIEGSTAYVYVADAEIKEDKVRIPYOISNKRNIKLTFIGHKGSEFT 660
Db 601 TPVDSVLFQKNIEDSEIAYIYNPDGEIGQIDKIKPISIDRPKILTFIGHGKDEFT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCNMFSYSIYABETYPKLLKIK 720
Db 661 DIPAGFDVDSLSSEIETILNLAADISPKYIEINLLGCNMFSYSINVEETYPKLLKIK 720
QY 721 DRYSELMPSTSDSITVSANQYVRINEEGRILDHSGKWKINKEISIKDISKEYISF 780
Db 721 DKISLMPSTSDSITVSANQYVRINEEGRILDHSGKWKINKEISIKDISKEYISF 780
QY 781 NPENKIIVKSKYLHELSTLQIRNANSDDILEKKVMTBCEINVASNIDRQIVGR 840
Db 781 NPENKIIVKSKYLHELSTLQIRNANSDDILEKKVMTBCEINVASNIDRQIVGR 840
QY 841 IEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHSPFISDICTENGPRIF 900
Db 841 IEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDSDHSPFISDICTENGPRIF 900
QY 901 INKETSPIETEKIFSEYATHISKEISNIKOTIFDVNGKLVKKNLDAAEVNTLN 960
Db 901 INKETSPIETEKIFSEYATHISKEISNIKOTIFDVNGKLVKKNLDAAEVNTLN 960
QY 961 SAFFIOSLIBYNTTKESLSNLVAMQVYAQLFSTGLNTITDASKVWELVSTALDETID 1020
Db 961 AAFIOSLIBYNTTKESLSNLVAMQVYAQLFSTGLNTITDASKVWELVSTALDETID 1020
```

## RESULT 4

```
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 636518
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCES/DOCKET NUMBER: OPHD-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-10

Query Match 78.7%; Score 4086; DB 4; Length 2366;
Best Local Similarity 77.5%; Pred. No. 5.1e-237;
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYKFRIOEDYVAIINALSEYHNMSSESVVEKYLKLDINLNTDNYL 60
Db 1 MSLVNRKQLEKQVYKFRIOEDYVAIINALSEYHNMSSESVVEKYLKLDINLNTDNYL 60
QY 61 NTYKSGRNLKALKEKFEYLTVTEVLELKNLSPVEKNLHFVIGGOINDTAINYNQWKD 120
Db 61 DTYKSGRNLKALKEKFEYLTVTEVLELKNLSPVEKNLHFVIGGOINDTAINYNQWKD 120
QY 121 VNSDVTYKVEYDSNAFLINTLKTIVESATNTTLESFRENLDNPDPEYKFKRMEIYY 180
Db 121 VNSDVTYKVEYDSNAFLINTLKTIVESATNTTLESFRENLDNPDPEYKFKRMEIYY 180
QY 181 DKQKHFDYKSOIBENPEFIIDNLIKTYLSNEYSKOLEALNKYTEESLNKITTANNNDI 240
Db 181 DKQKHFDYKSOIBENPEFIIDNLIKTYLSNEYSKOLEALNKYTEESLNKITTANNNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCYSDLVINOIKRYKILNDNLNPSINE 420
Db 361 SSGDMEASPLEVKIAFANNVINOALISLKDSCYSDLVINOIKRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKITYLVKGFAPDVRSTINLSPGQVYTCAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMELGKLVGVFPDVKTTINLSGPEAYAAQD 480
QY 481 LLMFKDGNSTNIHLPELNPFPKTKISQLTQEITSLWSFNQARAKSQFEYKGYE 540
Db 481 LLMFKEGSMNIHLIEADLNFEISKTNISQSTEQEASLWSFDDARAKAQFEYKRYE 540
```

541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTKRNEYIHYIYVLOQDKISYEASCNLFPSK 600  
541 GSGLEDNDLDFSONIVVDKEYLLEKISSLSRSGYIHYIYVLOQDKISYEACNLPFAK 600  
601 DYSSTLYOKNIEGSTAYYYVADAEIKEDYRIPYOISNKNRNLKLTFIGHGKSEFNT 660  
601 TPYDVLFOKNIEDSIAIYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNT 660  
661 DTFANLVDLSLSEIETILNADKADISPKYIEINLLGCNMFYSIYAEETYPGKLLKIK 720  
661 DIFAGFDVDSLSEIEAIDLAKEDISPKSIEINLLGCNMFYSINVEETYPGKLLKIK 720  
721 DRVSELMPSISQDSITVSANQYEVRIINEEGKEILDHSGKWNKESIIKDISSEYISF 780  
721 DKISELMPSISQDSIIIVSANQYEVRIINEEGKEILDHSGKWNKESIIKDISSEYISF 780  
781 NPKENKIIVKSKYLHELSTLLQEIIRNANSDDIDLEKKVMTCEINVASNIDRQIVEGR 840  
781 NPKENKITVKSKNLPESLTLQEIIRNANSDDIDLEKKVMTCEINVASNIDRQIVEGR 840  
841 IBEAKNLTSDSINYIKNEFKLIESISDLYDLKHQGLDDSHFISPEDISKTENGPRIRF 900  
841 IBEAKNLTSDSINYIKNEFKLIESISDALCDLKQNELEDHSHFISPEDISKTENGPRIRF 900  
901 INKGTGNSPIETEKEIFSEYATHISKEISNINOTIFDNNVNGKLVKKNLDAAEVNTLN 960  
901 INKGTGESIFVETKIFSEYANHITTEISKIGTIFDNNVNGKLVKKNLDTTAEVNTLN 960  
961 SAFFIOSLIEYNTKESLNSLVAMKVQYVQYAFSTGLTIITDASKVVLVSTALDETID 1020  
961 AAFPIOSLIEYNSKESLNSLVAMKVQYVQYAFSTGLTIITDAAKVVVLVSTALDETID 1020

RESULT 5  
US-10-011-366-10  
Sequence 10, Application US/10011366  
Patent No. 6573003  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
Kink, John A.  
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
DISEASE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/10/011,366  
FILING DATE: 16-No. 6573003-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FILING DATE: 23-OCT-1997  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingollia, Diane E.

REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-011-366-10

Query Match 78.7%; Score 4086; DB 4; Length 2366;  
Best Local Similarity 77.5%; Pred. No. 5.1e-237;  
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MNLVNAQLQKMYVYKFRQIEDEYVAIINLNALEBYEHNMSSSVVEKYKLKIDINLNTDYL 60  
DB 1 MSLVNRKQLEKMANVRFRTQDEYVAIILDALEBYEHNMSSSVVEKYKLKIDINLNTDYL 60  
QY 61 NYTKSGRNKALKKPKYLTMTVELKNNSLTPVEKNLHFVIGGQINDTAINYNQWKD 120  
DB 61 NYTKSGRNKALKKPKYLTMTVELKNNSLTPVEKNLHFVIGGQINDTAINYNQWKD 120  
QY 121 VNSDYTVKPYDSNAFLINTLAKTTVESATNTLTESFRENLDPEFDYKNFKRMEIY 180  
DB 121 VNSDYTVKPYDSNAFLINTLAKTTVESATNTLTESFRENLDPEFDYKNFKRMEIY 180  
QY 181 DKQHFIDYKSKQIEBENPEFIIDNIKTYSNEYSKDEALNKYIEESLKNITANNNDI 240  
DB 181 DKQFNFINYKAQREENPELLIDIVKTVLSNEYSKEIDELNTYIEESLKNITQNSGNDV 240  
QY 241 RNLEKFAEDLVRLYNQELVERWNLAASDIIRISMLKEDGGVLDVDPGLQPOLPKS 300  
DB 241 RNLEKFAEDLVRLYNQELVERWNLAASDIIRISMLKEDGGVLDVDPGLQPOLPKS 300  
QY 301 INKPSDITNTSWEMIKLEAIMKYKIPGYTSKNFMDLDEEVQSPESALSSKSKSEIF 360  
DB 301 INKPSDITNTSWEMIKLEAIMKYKIPGYTSKNFMDLDEEVQSPESALSSKSKSEIF 360  
QY 361 LPLDDIKVPLEVKIAFANNVINQALISLKDSCYSDLVINQIKRYKILNLPNSINE 420  
DB 361 LPLDDIKVPLEVKIAFANNVINQALISLKDSCYSDLVINQIKRYKILNLPNSINE 420  
QY 421 GTDFNTMTKIPFSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGGPVYTGAYD 480  
DB 421 GTDFNTMTKIPFSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGGPVYTGAYD 480  
QY 481 LLMFKDNTNTHLEPELRNPFPPKTKISOLTEQETISLWSFNQAPAKSQFEYKKGYPE 540  
DB 481 LLMFKDNTNTHLEPELRNPFPPKTKISOLTEQETISLWSFNQAPAKSQFEYKKGYPE 540  
QY 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTKRNEYIHYIYVLOQDKISYEASCNLFPSK 600  
DB 541 GSGLEDNDLDFSONIVVDKEYLLEKISSLSRSGYIHYIYVLOQDKISYEACNLPFAK 600  
QY 601 DYSSTLYOKNIEGSTAYYYVADAEIKEDYRIPYOISNKNRNLKLTFIGHGKSEFNT 660  
DB 601 TPYDVLFOKNIEDSIAIYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNT 660  
QY 661 DTFANLVDLSLSEIETILNADKADISPKYIEINLLGCNMFYSIYAEETYPGKLLKIK 720  
DB 661 DIFAGFDVDSLSEIEAIDLAKEDISPKSIEINLLGCNMFYSINVEETYPGKLLKIK 720  
QY 721 DRVSELMPSISQDSITVSANQYEVRIINEEGKEILDHSGKWNKESIIKDISSEYISF 780  
DB 721 DKISELMPSISQDSIIIVSANQYEVRIINEEGKEILDHSGKWNKESIIKDISSEYISF 780  
QY 781 NPKENKIIVKSKYLHELSTLLQEIIRNANSDDIDLEKKVMTCEINVASNIDRQIVEGR 840  
DB 781 NPKENKITVKSKNLPESLTLQEIIRNANSDDIDLEKKVMTCEINVASNIDRQIVEGR 840

```

QY 841 IBEAKNLTSDINVIKNEPKLIESISDSLYDLKHQGLDSDSHFISPEDISKTEGSPRIF 900
Db 841 IBEAKNLTSDINVIKNEPKLIESISDALCDLQOQNELEDSDHFIISFEDISEDEGFSIRF 900
QY 901 INKETGNSPIETEKEIFSEYATHISKEISNIKDTIFDNNVNGKLVKKVNLDAAEHVNTLN 960
Db 901 INKETGESIFVEKTIPISEYANHITEEISKIKGTIFDVTNGKLVKKVNLDTTHEVNTLN 960
QY 961 SAFFIOSLIEYNTKESLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 AAFPIOSLIEYNSKESLNSLVAMKVQVYAQLFSTGLNTITDAAKVVVELVSTALDETID 1020

RESULT 6
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHIE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

MOLECULE TYPE: protein

US-08-480-604A-6

```

Query Match 53.6%; Score 2783; DB 1; Length 2710;
Best Local Similarity 52.1%; Pred. No. 1.1e-158;
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

QY 1 MNLVNAQOLQKVVYKFRIOQDEYVAIILNAALEVEYHNMSESSVVEKYKLKDINLNTDNYL 60
Db 1 MSLISKEBELIKLAY-SIRPRENEYKTILTNDEYNKLTNNNNENKYLQKKLINESIDVFM 59
QY 61 NTKYKSGRNKALKKFKKEYLTMEVLELKNLSLTPPEKNLHFHWIGQINDTAINYNQMKD 120
Db 60 NKYKTSRRNALSNLKQOILKEVILIKNSNTPVEKNLHFHWIGEVSDIALEYIKQWAD 119
QY 121 VNSDYTVKPVYDSNAPLINTLKTTIVESATNNTLESFRENLDNPDYNNKFRKRMETIY 180
Db 120 INAEYNIKLWYDSEAFVNTLKAIVESSTTEALQLEBEEIQNPQDFNMKPYKKRMEFY 179
QY 181 DKQKHPIDYKSOIEENPEFIIDNIIKTYLSNEYSKDLEALNKYIEESLNKITTANNNDI 240
Db 180 DRQKRFINYKSIQINKPTVTIDDIKSHLSEYNDETVELSYTNSLRKINSNHGIDI 239
QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDILIRISMLKEDGGVYLDVILPGIQDLPKS 300
Db 240 RANSLFTEQELNIYSQELLNRGNLAAASDIIVLLALKNFGVYLDVMDLPGHSDLPKT 299
QY 301 INKPSITNTSWEMIKLEAIMKYKYPGYTSKNFDMDEEVQRFESALSKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYNNTYTNFDFKLDQQLKQKFLKIESKSEIF 359
QY 361 LPLDDIKVPLEVKIAPANNVINQALJSLKDSYCDLVINOIKRKYTLNLDNAPSINE 420
Db 360 SKLENLVSDLEIKIAPALGVSINOALISKQGSYLTNLVIEQVKRYQPLNQHLPABIS 419
QY 421 GTDFNTTWKIFSDKLASINEDNMFMKITNYLVKGFAPDVRSTINISGPGVYTCAYQD 480
Db 420 DNNFTDTTKIFHDSLNFNSATAENSMFLTKIAPYLVQVGFMPPEARSTISLSPGAYASAYD 479
QY 481 LLMFKDNSTNIHLLPELRNPEFPKTKISQLTQEBITSLWSNQARAKSQFEYKKGYPE 540
Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTQEBINSLSWSPDOASAKYQFEKYRDTYG 539
QY 541 GALGEDDNLDFQNTVLDKDY-VSKKILSS--MKTRNKEYIHVIQLOQDKISYASCNL 597
Db 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEEAGSKNYHYIIQLQGDDLSYATCNL 599
QY 598 FSKDPYSSILYQKNEGSEYAYYVAD--ABIKEDIKYRIPIYQISNKNINIKLTFIGHGK 655
Db 600 FSKNPKNSIIQORN--NESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGK 657
QY 656 SEFNTDTFANLVDVSLSEIETILMAKADISPKYIEINLLGCNMFPSYSIYAEETYPGKL 715
Db 658 DEFNTSEFARLSDVSLNSEISSFLDTIKLIDISPKQVEVNLLGCNMFPSYDFNVEETYPGKL 717
QY 716 LLKIKDRVSELMPSISQDSITVSANQYVRINEEKGREILDSHGKWINKEESIIDKISSK 775
Db 718 LLSIMDKITSTLDPVKNKSITIGANQYVRINSEGRKELLAHSGKWINKEEAIMSDLSK 777
QY 776 EYISFPNKENKIIVKSKYLHELSTLLQEIIRNANSDDIDLEKKVMLTECEINVASNIDRQ 835
Db 778 EYIFFDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFLNLLKNIIESIGDY 837
QY 836 IVEGRIEEAKNUTSDSINVIKNEPKLIESISDSLYDLKHQGLDSDSHFISPEDISKTEG 895
Db 838 IYEEKLEPVKNI IHNSIDDLDEFNLENVSDLEYELKKNLNDKYLISFEDISKONST 897
QY 896 FFIREFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNNVNGKLVKKVNLDAAE 955
Db 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQDHTSQ 957
QY 956 VNTLNSAFFIOSLIEYNTTTSKSLNSLVAMKVQVYAQLFSTGLNTITDASKVVELVSTAL 1015

```

Db 958 VNTLNAFFIQSLIDYSSNKVDLNDLSTSVKQVLAQFSTGLNTIYDSIQLVNLISNAV 1017

Qy 1016 DETID 1020  
Db 1018 NDTIN 1022

RESULT 7  
US-08-405-496A-6  
; Sequence 6, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPND-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2710 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-405-496A-6

Query Match 53.6%; Score 2783; DB 2; Length 2710;  
Best Local Similarity 52.1%; Pred. No. 1.1e-158;  
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

Qy 1 MHLVNAQLOKQVYKFRQEBYVAILNALBEYHNMSSESVVYKYLKLDINLTDNYL 60  
Db 1 MSLISKEELIKLAY-SIRPRENEYKTLTNLDEYNKLTNNENKYLQLKCLNESIDVFM 59  
Qy 61 NTYKSGRNKALKCKPEYLTMEVLELKNLTPVEKNLHPITWGGINDTAINYNOWKD 120  
Db 60 NKYTSRRNALSNLKKOILKEVILIKNSNTSPVEKNLHPVWIGVSDIALEYIKQWAD 119  
Qy 121 VNSDYTVKPYVDSNAPFLINTLTKKTIVESATNTNLFESFRENLDNPEFDYKNPKYKRMETIY 180

Db 120 INAEYNIKLYWYDEAFVNTLTKAIVESSTTEALQLEEEIQNPQFDNNKPKYKRMETIY 179  
Qy 181 DKQKHFDYKSKQIENPEFIIDNIITKYLSEYKOLEALNKYTEESLANKITANNNDI 240  
Db 180 DRQKFPINYKSKQINRPVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGIDI 239  
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDIIRISMLKEDGGVYLDVLDLPGIQDLPFKS 300  
Db 240 RANSLFTEQELLNIYSQELLNRLNLAASDI VRLALKNFGVYLDVLDVLMPLGHSDFLKT 299  
Qy 301 INKPSITNTSWEMIKLEIAIMKYKEYIPOYTSKNFMDLDEEVORFESALSSEKSKSEIF 360  
Db 300 ISRPSSIGLDRWEMIKLEIAIMKYKYINNTYTSNFDFKLDQQLKDLKFLKIESKSEKSEIF 359  
Qy 361 LPLDDIKVSPLEVKIAFANNVINQALISLKSYSGLVINOIKRYKILNDNLNPSINE 420  
Db 360 SKLENLVSDLEIKIAFALGVSINOALISKQSYLTNLVIEQVKRYQFNLQHLNPAIES 419  
Qy 421 GTDFNTWKIFSDKLASISNEDNMFMKITNYLVKGPAPDVRSRTINLSGPGVYTTGAYOD 480  
Db 420 DNNFTDTTKIPHDSLFSNATAENSMFLTKIAPYLQVGFMPPEARSTISLSGPGAYASAYD 479  
Qy 481 LLMFKDNSTNHLLEPELNPEFPKTKISQLTQEISLWSFNQARAKSQPEYKKGYPE 540  
Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTQEISLWSFDQASAKYQFEKYVRDYTG 539  
Qy 541 GALGEDDNLDFQAQNTVLKQY-VSKILSS--MKTRNKEYIHVIQLOGDKISYEASCNL 597  
Db 540 GSLSEDNGVDNFKNKTALDNKLNKIPSNVVEAGSKNYHYIIQLQGGDISYEATCNL 599  
Qy 598 FSKDPYSSILYQKIEGSETAYYYVAD--ABIKEIDKYRIPQYISNKNKIKLTFIGHCK 655  
Db 600 FSKNPKNSIIQORNM--NESAKSYFLSDDGESILELNKYRIERLCKNKKVKVTFIGHCK 657  
Qy 656 SEFNTDTTANLVDVSLSEIETILNLAKADISPKYIEINLLGCMFMSYSIYAEETYPGKL 715  
Db 658 DEFNTSEFARLSVDSLSENEISSFLDTIKLIDISPKNVEVNLGCMFMSYDFNVEETYPGKL 717  
Qy 716 LLKIKDRVSELMPSISQDSITVSANOYEVRINEEGKREILDHSGKWINKKEESIIDKISSK 775  
Db 718 LLSIMDKITSLPVDNKNKSITIGANQYEVIRINSEGRKELLASHGKWINKKEEAIMSLSK 777  
Qy 776 EYISPNKRNKIIVKSKYLHELSTLLQEIIRNNANSDDIDLEKKVMLTECEINVASNIDRQ 835  
Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSDTKFILNLLKLAIESIGDY 837  
Qy 836 IVEGRIBRANLTSDSINYIKNEPKLIESISLVDLKHQGLDDSHFISFEDISKTENG 895  
Db 838 IYVEKLEPVKNIHNSIDDLIDEPNLENVSDLEYELKLNLDKYLISFEDISKNST 897  
Qy 896 FRIRINKETGNSIFETEKEIFSEVATHISKEISNIKDTIPDNVNGKLAVKKNVLDAAHE 955  
Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLDNLDIDHTSQ 957  
Qy 956 VNTLNAFFIQSLIDYSSNKVDLNDLSTSVKQVLAQFSTGLNTIYDSIQLVNLISNAV 1015  
Db 958 VNTLNAFFIQSLIDYSSNKVDLNDLSTSVKQVLAQFSTGLNTIYDSIQLVNLISNAV 1017  
Qy 1016 DETID 1020  
Db 1018 NDTIN 1022

RESULT 8  
US-08-915-136-6  
; Sequence 6, Application US/08915136  
; Patent No. 6290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,136

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/480,604

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-915-136-6

Query Match 53.6%; Score 2783; DB 3; Length 2710;

Best Local Similarity 52.1%; Pred. No. 1.1e-158;

Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

QY 1 MNLVNAQLOKMYVVFRIQEDYVAILNALAEYHNMSSESVVEKYLKLDINNLVDNYL 60

DB 1 MSLISKEELIKLAY-SIRPRENEYKILTLDYKLTNNNNENKYLQKKNESIDVFM 59

QY 61 NTYKSGRNKALKFKREYLTMEVLELKNNSLTPEKRLHFIMIGGOINDTAIYNQWKD 120

DB 60 NKYTSRRNALNKKDILKEVILKNSTSPVKRLHFWIGGEVSDIALEYIKQWAD 119

QY 121 VNSDYTVKPYVDSNAFLINTLKKTVESATNTNLFRENLNDPEFDYNNFKYKRMELIY 180

DB 120 INAEYNIKLYDSEAFVNLTKKAIVESSTTEALQLLEEBIQNPQFDNMKFYKRMELIY 179

QY 181 DKQGHFDVYKQIEENPEFIIDNIITKYLNSYKDLKALNKYIEESLNKITANNNDI 240

DB 180 DROGRFINYKQINQKPTVTIDDIKSHLVSNRNDVLESYRTNSLRKINSNHGDI 239

QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQPLFKS 300

DB 240 RANSLFTEQELNLYSQELLNRNLAASDIVRLALKNFGVGLVDVMDPLGIHSDLPKT 299

QY 301 INKPSITNTSWEMIKLEAIKKYKEYIPGYTSKNFMDLDEEVORFESALSXSOKSEIF 360

DB 300 ISRPSSIGLDRWEMIKLEAIKKYKEYINNYTSSENFDKLDQDQKQKLFIESKSEKSEIF 359

QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLSDSYSDLVINOIKNRYKILNDNLPSINE 420

DB 360 SKLENLVSDLEIKIAPALGVSINOALISKQSYLTNLVIEQVKRYQFLNQHLPATBS 419

QY 421 GTDFNTTKIFSDKLASISNEDNMFMKITNLYKVGAPDVRSTINISGPGVYTGAYOD 480

DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVGFEMPEARSTISLSGPGAYASAYD 479

QY 481 LLMPKDNSTNIHLLEPELRNPEFPKTKISOLTEQBITSLWSFNQARAKSOFEEYKGYFE 540

DB 480 FINQENTIEKTLKASDLIEFKFPENNLSQLTEQBSINLSWSPDQASAKYQFEKYVRYDYG 539

QY 541 GALGEDDNLDFQANTVLDKDY-VSKKILSS--MKTRNKEYIHYIYVLOQGDKISYEASCNL 597

DB 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKNYVHYIQLQGGDISYEATCNL 599

QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKSIDKRYPIQYISNKNIKLTFIGHOK 655

DB 600 FSKNPKNSIIQORNM--NESAKSYFLSDDDGESILELNKYRIPELKNKEKVKVTFIGHOK 657

QY 656 SEFNTDTFANLVDVDSLSSEIETILNLAADISPKYIEINLLGCNMFYSIYAEETYPGKL 715

DB 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLIDISPKNVEVNLGCNMFYDFNVEETYPGKL 717

QY 716 LLKIKDRVSELMPSISODSITVSANQYEVINEEGKREITLDHSGKWINKESIIKDISK 775

DB 718 LLSIMDKITSTLPDVNKSITIGANQYEVIRINSEKRELLAHSKWKINKERAIMSDLSK 777

QY 776 EYISFPKENKIIIVSKYLHELSTLLQEIIRNNANSDDIDLEKKVMLTECEINVASNIDRQ 835

DB 778 EYIFFDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSPDTKFLNNLKLNISSIGDY 837

QY 836 IVEGRIEAKNTSDSINVIKNEPKLIESISLSDVLDKHQGLDSDSHSFEDISKSTENG 895

DB 838 IYYEKLEPVKNIIHNSIDDLIDEFNLEENVSDELVELKLNLDLDEKYLISFEDISKSNST 897

QY 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNICKDTIFDNVNGKLVKKNLDAAEH 955

DB 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNNLLDNIQLDHTSQ 957

QY 956 VNTLSAFFIQSLIEYNTTKESLSNLSVAMKYOVYQAFLPSTGLNTITDASKVELVSTAL 1015

DB 958 VNTLNAAPFIQSLIDYSSNKNVDNLSTSVKQVLAQLPSTGLNTIYDSIQLVNLISNAV 1017

QY 1016 DETID 1020

DB 1018 NDTIN 1022

RESULT 9

US-08-957-310-6

; Sequence 6, Application US/08957310

; Patent No. 6365158

; GENERAL INFORMATION:

; APPLICANT: Williams, James A.

; APPLICANT: Kink, John A.

; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES

; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

; TITLE OF INVENTION: DISEASE

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medlen & Carroll

; STREET: 220 Montgomery Street, Suite 2200

; CITY: San Francisco

; STATE: California

; COUNTRY: United States of America

[illegible]

/ FILING DATE: 02-DEC-1993  
/ APPLICATION NUMBER: US 07/985,321  
/ FILING DATE: 04-DEC-1992  
/ APPLICATION NUMBER: US 07/429,791  
/ FILING DATE: 31-OCT-1989  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Ingolia, Diane E.  
/ REGISTRATION NUMBER: 40,027  
/ REFERENCE/DOCKET NUMBER: OPHD-01121  
/ TELEPHONE: (415) 705-8410  
/ TELEFAX: (415) 397-8338  
/ INFORMATION FOR SEQ ID NO: 6:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 2710 amino acids  
/ TYPE: amino acid  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-011-366-6

Query Match 53.6%; Score 2783; DB 4; Length 2710;  
Best Local Similarity 52.1%; Pred. No. 1.1e-158;  
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;

QY 1 MNLVKAQLQKVVYKFRIOEDYVAI LNALBEYHNMSSESVVEKY LKLKLDNNLT DNYL 60  
Db 1 MSLISEELIKLAY-SIRPRENEYKTLTNDYENKLTNNENKYLQKKLINESIDVPM 59  
QY 61 NTYKSGRNKALKKKPKXYLTMEYLELKNNSLTPVEKNLHPFIWIGQINDTAIYINQWKD 120  
Db 60 NKYKTSRRNALSNLKKDKILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD 119  
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTLESFRENLDNDPEPDYKNFKRKRMEIYY 180  
Db 120 INAEYNIKLWYDEAFVNLTKKAIYVESSTTEALQLEBEEIQNPQPDNNKFKYKKRMEFIY 179  
QY 181 DKQKHFIDYVKSQIENPEFIIDNIITKYLSEYKDLALNKYIEESLNKKITANNNDI 240  
Db 180 DRQKRFINYKQINQKPTVPTIDDIILKSHLVSEYNRDETVLESYRNSLRKINSNHGIDI 239  
QY 241 RNLEKFAEDVLVRLNQELVERNLAAASDILRISMLKEDGGYVLDVILPGIQDPLFKS 300  
Db 240 RANSLFTEQELLNIYSQELLNRGNLAAASDIVRLALKNFGVYLDVDMPLGHSDFLKT 299  
QY 301 INKPSDITWSEMIKLEAIMKYEYIPGYTSKNFMDLDEEVORFESALSSEKSKSEIF 360  
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQKDNPKFLIESKSEKSEIF 359  
QY 361 LPDDLIKVPLEVKIAFANNSVINOALISIKDSCYSDLVINOIKNRYKILNDNLPSINE 420  
Db 360 SKLENLVSDLEIKIAPALGVSINOALISKQGSYLTNLVIEQVKRYQFLNQHNLPAIES 419  
QY 421 GTDFNTMTKIFSDKLASIGNEDNMFMKITNLYKVGFAFPVRSTINLSGPGVYTTGAYOD 480  
Db 420 DNNFTDTTKIFHDSLSNATAENSMLTKIAPVLOVGFPEARSTISLSGPGAYASAYD 479  
QY 481 LMFKNSTNIHLLPELRFNFPKTKISOLTEQETISLWSFNQARAKSQPEYKKGYPE 540  
Db 480 FINLOENTIEKTLKASDLIEFKPEPNLSQLEINLSWSPDQASAKYQFEKYVRDYTG 539  
QY 541 GALGEDNDLPFAONTVLDDXDY-VSKKILGS--MKTRNKEYIHVIVOLQGDKISYEASCNL 597  
Db 540 GSLSENGVDVFNKATLDKRYLLNKKIPSNVVEAGSKYVHYIILQGGDDISYEATCNL 599  
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEIDKRYIPYQISNKRNIKLTPFGHGK 655  
Db 600 FSKNPKNSIIIOIRM--NESAKSYFLSDGESILELNKYRIPERLNKKNKVKVTFIGHGK 657  
QY 656 SEFTNTDFANLVDVSLSSSETETILNLAADISPKYTEINLLGCMFSEIYAEETYPGKL 715  
Db 658 DEFNTSEFARLSVDSLSEISSEFLDTIKLDISPKNVENVLLGCMFSEYDFNVEETYPGKL 717

QY 716 LLKIKDRVSELMPSISODSITVSANQYEVIRINEEGREKILDHSGKWINKEESI IKDISK 775  
Db 718 LLSIMDKITSTLPDVNKNISITIGANQYEVIRINEEGREKELLAHSGKWINKEEAIMSDLSSK 777  
QY 776 EYISNPKNKIYVSKYLHELSTLLQEIQRNANSSDIDLEKKVMLTCEINVASNIDRQ 835  
Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFTLNNLKLNISSIGDY 837  
QY 836 IVEGRIBEAKNLTSIDNINIKNEFKLIESISDSLYDLKHQNGLDDSHFISPEDISKTEG 895  
Db 838 IYVEKLEPVKNIHNSIDDLIDEFNLEENVSELYELKLNLDDEKYLISFEDISKNNST 897  
QY 896 FRIRFINKETGNSIFETETEKEIFSEYATHISKEISNIDKTIPTDNVNGKLVKKNVDAHA 955  
Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNILQDHTSQ 957  
QY 956 VNTLSAFFIOSLIENYNTTKSLSNLSVAMKVQVYQAQLFSTGLNTITDASKVVELVSTAL 1015  
Db 958 VNTLNAAFFIQSLIDYSSNKKVDLNDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLSINAV 1017  
QY 1016 DETID 1020  
Db 1018 NDTIN 1022

RESULT 11  
US-09-453-702B-257  
; Sequence 257, Application US/09453702B  
; Patent No. 6365723  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; Burland, Valerie  
; Berna, Nicole T.  
; Plunkett, Guy  
; Welch, Rod  
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/453,702B  
; FILING DATE: 03-Dec-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/110,955  
; FILING DATE: 04-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 251-5000  
; TELEFAX: (608) 251-9166  
; INFORMATION FOR SEQ ID NO: 257:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3169 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 257:  
US-09-453-702B-257



Query Match 7.9%; Score 411.5; DB 4; Length 3169;  
Best Local Similarity 20.9%; Pred. No. 3.6e-16;  
Matches 257; Conservative 195; Mismatches 453; Indels 323; Gaps 54;

QY 14 YKPRI---QEDYVAILNAL---EYHNMSESVVEKYLKIDINLNDYNTYNTYKSGR 68  
DB 160 YIKIRKTRGAEDQTTITQSLINELLNGVDRTTI--PPQKISELNDIHSYENMQIKNSR 218  
QY 69 N--KALKKPKFYLTWEVLELKNS--LTPVEKNLHFTWIGQ 106  
DB 219 KGIELVQKQELLSSLIINDKNGKQSDNASKIINLLGIEYQSHKVDIEPFIHAWVAGA 278  
QY 107 INDTAINYNQKVDNSYTVKXVDSNAF-----LINTLAKT----- 144  
DB 279 PDNTFSVITAFNTYKDYTLWIDPNAFGAAPKSGILKNTAMVAINMLRRTTNPHLAE 338  
QY 145 -----IVESATNTLE--SFRENLDPEFY-----NKFYRKMELIYDQKIFI 187  
DB 339 EMNEVLKIQNTONETIEFKETRERLKELENRYKSLTSTKPKNFVFFLESMIGHQDNFY 398  
QY 108 DYKSOIEENPEFI-----IDNIK--TVLSNEYSKQLEALNKYIEESLNKITTANN-- 238  
DB 399 TYCISNGISNTDIDSLDFLTWVVKLSEVQDFKSTVEKNKRDIDLKNTISQKFGDRF 458  
QY 239 ---DIRNLEKFADEDLVRLYNQELVERNMLAAASDILRISMLKEDGVYLDVILPGIQP 295  
DB 459 QLRDINTLESFKPDQPPYQEMLLRWYAAASQVRINILKEGGIYDTDILPAYSD 518  
QY 296 DLFPKINKPDSITNTSWEMIKLEAMKYK--BYIPG--YTSQVMDLBEVORSFESALS 351  
DB 519 KVSQIINE--KSDKRFPPEDLKLRIISELSILKGEKYSIKH--DGLDDETLNQLNNIL- 575  
QY 352 SKSDSEIFLPDDIKVSPLEVKIAPANNV-----NOALISLKDSVC 395  
DB 576 SIEK-----LTDDY--FRVETKVRDTPKIFRQKVTENTWIRGNMFLMTHGSKC 630  
QY 396 SLDVINQIKRYKIILNDNLNPSINEGTDFNTMKIFSDKLASISNE-----DNM 444  
DB 631 IDPILSGQKKQYLEL--QRIRDNISYNLFPYTT-----EDLSLNNVAIGGIPAKKYLEHG 684  
QY 445 MPMIKITVYKVGPPADVSTINLSGP----- 471  
DB 685 LF-----SEYRODGTPIYVYSTLNISGPDIMRMQMKYKSLGRIGEVHIKONKLSDVNPL 740  
QY 472 GYVTGAYQDLMF---KONSTNIHLEPE-----LRNPEFPKTKISQLT-----BOEITS 518  
DB 741 GYASNKNKDSFNWLPVSGINDITPDDESSWAVRNNNDINKILFEKINCHVPEKLPST 800  
QY 519 LMSFNQARAKSQFEYKKGYPGALGEDDNDLDFAQNTVLDKDYVSKKILSSMKTNRKEYI 578  
DB 801 LY-----YEIDSRSPFOGW---DNKSIKHVTEINKDLI--KOINLLLTSSNDV 844  
QY 579 HVIVLQGGKISYEASCNLFKDPYSSILYQKNIEGSETAYYYYVADAEIKIDKRIPIY 638  
DB 845 KLLIKL--DRELYAISKI-----DNPLALSIRTLQLLANYVTNTFEPENTINFY 896  
QY 639 QISNKRN-----IKLTFIGHCKSEBNTDTFANLDVS-----LSSEIET 677  
DB 897 DYRKQDDLLSAIKL-----FSRNDADTKIIWVNSVMEKNVFLREVISC 942  
QY 678 IINLAKADISPKYIEINLLGCMNFSYIYAEBETYPKGLLLKIKDRVSELMPSISQDSI-- 735  
DB 943 VLRSKKVD--SYINEN-----KKNLSKEDA--GALRDYAKLAKMKELFSLMDDGGYCK 990  
QY 736 TVSANQYEVRIINEEGREILDSGKWKINKEESIIKDISEKEYISNP-----K 783  
DB 991 IITTNAY-----IKERDKL-----SGIYNIENSIISGHESFDIIRSNQHEWGLSTVEOPK 1042  
QY 784 ENKIIVKSHLSTLLOEIRNNANSSDIDLEKKVMLTECHINVAS-----NIDRQIVE 838  
DB 1043 KPEFVVKSE--LSSAKSIFDDIKK--YTDPETKRNVLVHQLDSIDKRIAFLDIDSHYAVP 1100  
QY 839 GRIEBAKNLTSINSYKNEFKLIESI-----SDSLYDLKHQNGL 878

DB 1101 GSLLEKLQLS---GYVPSDINIITAEYLLASGVSHGYVVPAPSDKLELLRRHTK 1156  
QY 879 DDSHFISPIDISKTEGPRIRINKETGNSI---PIETEKEIFPSYATHISKEISNIKDT 935  
DB 1157 SNSEWI---EKITP---YVYDILSDNVSNNVLRPPUSEQKILNDIKLEISKVS---EQ 1207  
QY 936 IFDYNVNGKLVKKVNLDAAEVNTLNSAFFIOSLIEYNTTK-----ESLSNLSV----- 983  
DB 1208 YPMKLTQEKSSVIGIKYSVDPRYNENLFLSLPINQNLTLPMYRYFEMLYDIHIGIEN 1267  
QY 984 -AMKVQVYAQLFSTGLNTITDASKVVEL 1010  
DB 1268 KANREFIYSKFFSSLMLDLFLINDERVLNL 1295

RESULT 12  
US-09-417-485D-6  
; Sequence 6, Application US/09417485D  
; Patent No. 6541202  
; GENERAL INFORMATION:  
; APPLICANT: Long, David M.  
; APPLICANT: Metz, Anneke M.  
; APPLICANT: Love, Ruschelle A.  
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes  
; FILE REFERENCE: 47714-5009-US  
; CURRENT APPLICATION NUMBER: US/09/417,485D  
; CURRENT FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2184  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (330)..(335)  
; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;  
; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.  
US-09-417-485D-6

Query Match 5.0%; Score 261.5; DB 4; Length 2184;  
Best Local Similarity 19.7%; Pred. No. 2.2e-07;  
Matches 252; Conservative 193; Mismatches 442; Indels 395; Gaps 61;

QY 10 QKVYVVKRIQDEVAITNALAEY--HNMSES---SVVEKYLKLDI-----NNLTD 57  
DB 572 KKCIPKILGSKHNFKIFLKNVKKFLFLFNFKESFSLNQVMKNIKVKNIFQKKISKYINKN 631  
QY 58 NVL--NTYKSGRNKAL--KPKFELYT---MEVLELKNLSLTPVEKNLHFTWIGQIND 109  
DB 632 RILLKNIIFDNNYENKILHRNNKEIITNDINIKYKNKDNL-----NNSP-----KKT 681  
QY 110 TAIN-----YINQWKVN-----SDYTVKPVVDSNAFLINTLUKKTIVESATNTL-- 154  
DB 682 TLENKLRKYFNKIKKINIAIQKRLMNRILYFLF--NYFIMPLIRRRFFLTQKSBOTLHK 739  
QY 155 -----ESPR-----ENLNDPEFY--NKP-YRKRMEIYDKQKH 185  
DB 740 TIFDRKIWNHFTKISNFCYHQIFRNKGLKGRNEPKMDYVQNMFNVKKGSKI--KTNK 797  
QY 186 FIDYVKSQIEENPEFIIDNIKTYSNEYSKQLEAL---NKYIEESLNKITTANN---GN 238  
DB 798 YIFIKMKKKSTKCNKNNKFKKCPKCKKKKNLYNITRHNNIFIKKDKKKSTNNLNK 857  
QY 239 DIRNLEKPAD--EDLVRLYNQEL-----VERNMLAAASDILRISMLKEDGVYLDVILPGI 293  
DB 858 SIDNLYLKLKEINKKSVRPYIKKFFYKIKKKYFALKKKYIHRMMAKEE----- 904  
QY 294 QPDLEKSNKPDSTINTSWEMIKLEAMKY-----KEYIPGYTS-----KN 334  
DB 905 -----KS-----NIKLERAFKHFIFPAQBEKHLKYPFSSHPFQNRKINYGR 946



QY 794 LHELSTLQEIIRNAN-----SSDIDLEKKVMLTECEINVASNIDROI VEGRIEAEKVL-T 848  
DB 7459 LNSNMQLQIIVNNDTVKQNSDF-----INEDSS-QQDAYNHAIAQAADLIT 7505  
QY 849 SDSINYIKNEP-KLIESISDSLYDLKHQGLDDSHFISPIDISKTBNGPRIRPINKETG- 906  
DB 7506 AHTPTMDKQIQDAIENIKQALNDLHGSNKLSB-----DKKEASBQLQNLNSLTNG 7556  
QY 907 -----NSIF-IETEKEIESEVATHISKEISN-----IKDTIFDNVNGKLVKKNLDAAE 955  
DB 7557 QKDTILNHIFSAPTRSQVEKTIAS--AKQLNNTMKALRDSIADNNEILQSSKYFNDSBQ 7614  
QY 956 VNTLSAF-----FIQSLIEYNTTKESLSNLSVAMKV---QVYAOLF 994  
DB 7615 QYANQAVNKAANIINDQTPVMADEIQSVL--NEVQTKONLHGDKLANDKTDQAAT 7672  
QY 995 STGLNITTDASK 1006  
DB 7673 LNALNYLNAQR 7684

RESULT 14

US-09-091-501B-10  
; Sequence 10, Application US/09091501B  
; Patent No. 6518413  
; GENERAL INFORMATION:  
; APPLICANT: Tinsley, Jonathon M  
; APPLICANT: Davies, Kay E  
; TITLE OF INVENTION: Utrrophin gene expression  
; FILE REFERENCE: 620-42  
; CURRENT APPLICATION NUMBER: US/09/091,501B  
; CURRENT FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: PCT/GB96/03156  
; PRIOR FILING DATE: 1996-12-19  
; PRIOR APPLICATION NUMBER: GB 9525962.8  
; PRIOR FILING DATE: 1995-12-19  
; PRIOR APPLICATION NUMBER: GB 9615797.9  
; PRIOR FILING DATE: 1996-07-26  
; PRIOR APPLICATION NUMBER: GB 9622174.2  
; PRIOR FILING DATE: 1996-10-24  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 3433  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; NAME/KEY: misc feature  
; LOCATION: (239) ... (250)  
; OTHER INFORMATION: Description of Artificial Sequence: Full length  
; OTHER INFORMATION: utrophin construct; Xaa = unknown  
US-09-091-501B-10

Query Match 4.6%; Score 237.5; DB 4; Length 3433;  
Best Local Similarity 18.8%; Pred. No. 1.1e-05;  
Matches 227; Conservative 212; Mismatches 450; Indels 321; Gaps 57;  
QY 27 ILNAAEYHNMSSESVVEKYLLKDINNLTNDYNTYKSGRNKALKKFKYLTMEVL-B 85  
DB 699 LLNLWIKWTAIQTTEIKETMKMQDTSEMKK-LKALEKEQREIRIPRADELNGTGQILVE 757  
QY 86 LKNSLTPVEKNLHPTWIGGQINDTAINYNQKVN-----DYTVKFVYDSNAP----- 136  
DB 758 QMGKEGLPTE-----EIKNVLEKVSSEKWNVSQHLDELKRIQLQEDINAYFKQL 807  
QY 137 -----LINT-----LKKTVESATNTLTSPRE-----NL----- 161  
DB 808 DELEKVIKTEKWEVKWKTSTSESRRQSLDKSCQRELTNLLGLHPKIEWARSASCALMS 867  
QY 162 --NDPEFDYKNP-----YKRMEEIYDKQKHFIDYKSO-----IENPEFIIDNI 205  
DB 868 QPSAPDFVQRGDFSLGRYQAVQEAVEDRQOHLNELKGPQGHAYLETTLTKLKDVLNDS 927

QY 206 IKTYLSNEYSKDEALNKYIEB--SLNKITANN-----GNDIRNLEKFADEDLVRLYN 256  
DB 928 NKAQVSLVNLNDLAKVEKALQEKTKTLDLEILENQPALHKLAEETKALEKVPDVEKLYK 987  
QY 257 QSLVE---RWL---LAAASDI-----LRISMLKEDGCV-----YLDVDLPGIQDPLFKS 300  
DB 988 QEFDDVQGWKVKLVLSKDLHLLEBEIALTLRAFEADSTVIEKWM-----GVKDFLMQ 1042  
QY 301 -----INKPDSITNTSWEMIKLEAIMK-----YKEIYP-----GYT 331  
DB 1043 QAAQGDAGLQRLQDOCSAFVNEIETISSLKNKMEIETNLSRSGPVAGIKTWQTRLGY 1102  
QY 332 SKNFMLDEEV-----QRSFESALSSKSKSEIFPLDDDIKVSPLYEVKIAFANNVIN 384  
DB 1103 OTQLEKLSKEIATQKSRLSBSQKAAKLLKDLAEQEMWTAQAEVEELDERFEVKSPEELE 1162  
QY 385 QALISLKOSYCSDLVINQIKRYKILNDNLNPSINEGTFDTMTKIFSKLASISNEDNM 444  
DB 1163 SAVEEMKRAK-EDVLQKEV--RVKILKONI-----KLLAAKVPSSGGQELTS 1205  
QY 445 MFWIKITNYL-----KVGFPDVRSTINLSGPGVYTGAYQDILLMPKD-NSTNIHLE 495  
DB 1206 ELNVVLENYQLLCNRIRGKCHTLEEVMS-----CWEILHYLDLETTWLTLE 1253  
QY 496 PELRNFPKPKTISOLTEQEIITSLMSF--NOARAKSQPEEYKKGYPGEGALGEDDNDLPAQ 553  
DB 1254 ERMKSTEVLPEKTDVNE-ALESLESVLHPADNTQIRELQTLIDG----- 1301  
QY 554 NTVLDKDYVSKKILSSMTKNEYIHYIVQLQDKIYSYBASCNLFSPKDPYSILYOKNIE 613  
DB 1302 --ILD-DIIESEK-LEAFNSRYEDLSHL---ASSKQISLEKQLQVLRRETQMLQVLQESLG 1354  
QY 614 GSETAYYYVADAEIKEDIKYRIPQIS-----NKNIKLTPFGCHK 655  
DB 1355 ELDKQLTTLTLD---RIDAFQVQPEAQKIQAIEISAHELTLEELRNMSQPLT----- 1404  
QY 656 SPFNFTDTPANLDVDSLSSEIETILNLAKADISPKYIEINLLGCNMFYSIYABETYPGKL 715  
DB 1405 SPESRTARGSQMDVLQRLREVSTYKFLQFQKPAFQRMLODKRVLDGVKAE----- 1457  
QY 716 LLKIKDRVSELMPSISQDSITVSANOQYEVRIINEBGRKREILDHSGKWINKEESI-K- 772  
DB 1458 -LHVLD-VKDDPDPVIQTHLDCMKLYKTL--SEVKLEV-----ETVIKTGRHI 1502  
QY 773 SSKEYISFNPKNKIIVKSKYLHELSTLLOETRNANSSDIDLEKKVMLTECEINVASNI 832  
DB 1503 VQKQQT-D-NPKG-----MDEQLTSLKVLNDILGAQVTEGQDLERASQLARKMKKEAASL 1556  
QY 833 DRQIVEGRJEEAKNLTSD-----SINYIKNEFKLIE-----SISDSLYDLKHQNG 877  
DB 1557 SEWLSATETELVQKSTSEGLDGLDTEISWAKNVLDLEKRAKADLNTITESSAAL--QNL 1614  
QY 878 LDDSHFISPIDISKTBNGF-RIRFINKETGNSIF-IETEKIFSEYATHISKEISNIKOT 935  
DB 1615 IEGSBELERLCVLNAGWSRVRTWEDWCNTLMNHQNLQEIFDGNVAHISTWLYQ-ABA 1673  
QY 936 IFDNVNGK-----LVKKV--NLDAH--EVTNLSAFFIOS-----LIEY 971  
DB 1674 LLDEIEKPTSQEEIVKELVSELDDANLQVENVRDQALILMNARGSSRELVEPPELAE 1733  
QY 972 NTKTESLS-----NLSVAM---KVQVYAOLFSTGL-----NTITDASKVVE--L 1010  
DB 1734 NNFVKVSHIISAKLLIAQEPYQCLVTTFETGVPFSDLEKLENDIENMLKPFVEKHL 1793  
QY 1011 VSTALDETID 1020  
DB 1794 ESSDEDEKMD 1803

RESULT 15  
US-08-353-700-1  
; Sequence 1, Application US/08353700



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:31:53 ; Search time 24.375 Seconds  
(without alignments)  
7187.047 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1020  
Perfect score: 5192  
Sequence: 1 MNLVNAQLQKVMVYKFRIQ.....ITDASKVELVSTALDETID 1020

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4086	78.7	2366	15	US-10-011-366-10
2	2783	53.6	2710	15	US-10-011-366-6
3	411.5	7.9	3169	15	US-10-114-170-257
4	261.5	5.0	2184	12	US-10-304-095-6
5	256	4.9	1948	12	US-10-032-585-7611
6	249	4.8	1639	15	US-10-087-484-10
7	246	4.7	1881	12	US-10-032-585-7646
8	224	4.3	1270	12	US-10-032-585-7127
9	221	4.3	1805	11	US-09-820-843A-73
10	218	4.2	2285	10	US-09-932-183A-2
11	213	4.1	1979	15	US-10-205-823-419
12	212	4.1	1610	15	US-10-155-533-9
13	209.5	4.0	993	9	US-09-815-242-5809
14	209.5	4.0	1002	9	US-09-815-242-12899
15	209.5	4.0	1002	9	US-09-815-242-13158

16	209.5	4.0	1240	12	US-10-032-585-7366	Sequence 7366, Ap
17	208.5	4.0	1786	10	US-09-742-096-3	Sequence 3, Appli
18	207	4.0	996	9	US-09-815-242-5251	Sequence 5251, Ap
19	207	4.0	1009	9	US-09-815-242-12141	Sequence 12141, A
20	205.5	4.0	696	12	US-10-032-585-7808	Sequence 7808, Ap
21	204	3.9	1171	12	US-10-032-585-7519	Sequence 7519, Ap
22	204	3.9	3878	12	US-10-080-608A-11	Sequence 11, Appl
23	202.5	3.9	1169	12	US-10-241-596-20	Sequence 20, Appl
24	202	3.9	872	10	US-09-843-676-8	Sequence 8, Appli
25	202	3.9	872	10	US-09-843-676-54	Sequence 54, Appli
26	202	3.9	872	10	US-09-766-253-8	Sequence 8, Appli
27	202	3.9	872	10	US-09-766-253-54	Sequence 54, Appli
28	202	3.9	872	11	US-09-438-486-8	Sequence 8, Appli
29	202	3.9	872	11	US-09-438-486-54	Sequence 54, Appli
30	202	3.9	872	15	US-10-053-758-8	Sequence 8, Appli
31	202	3.9	872	15	US-10-053-758-54	Sequence 54, Appli
32	202	3.9	872	15	US-10-054-295-8	Sequence 8, Appli
33	202	3.9	872	15	US-10-054-295-54	Sequence 54, Appli
34	202	3.9	872	15	US-10-054-611-8	Sequence 8, Appli
35	202	3.9	872	15	US-10-054-611-54	Sequence 54, Appli
36	201.5	3.9	1086	10	US-09-924-154-15	Sequence 15, Appl
37	200.5	3.9	3158	9	US-09-815-242-12611	Sequence 12611, A
38	199.5	3.8	2871	15	US-10-146-473-41	Sequence 41, Appl
39	199	3.8	3899	15	US-10-171-311-4	Sequence 4, Appli
40	199	3.8	3917	15	US-10-171-311-8	Sequence 8, Appli
41	198.5	3.8	1978	12	US-10-094-466-64	Sequence 64, Appl
42	197.5	3.8	2119	12	US-09-769-744A-28	Sequence 28, Appl
43	197	3.8	1579	10	US-09-801-368-368	Sequence 368, App
44	197	3.8	3907	15	US-10-171-311-2	Sequence 2, Appli
45	197	3.8	3925	15	US-10-171-311-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-10-011-366-10  
; Sequence 10, Application US/10011366  
; Publication No. US20030054493A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; Kink, John A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/011,366  
; FILING DATE: 16-NO. US20030054493A1-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,310  
; FILING DATE: 23-OCT-1997  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-011-366-10

Query Match 78.7%; Score 4086; DB 15; Length 2366;  
Best Local Similarity 77.5%; Pred. No. 2.5e-248;  
Matches 790; Conservative 105; Mismatches 125; Indels 0; Gaps 0;

QY 1 MLVYKALQKVVYKFRQEDYVAIALNABEYHNMSSESVVEKYKLKIDINLNTDYL 60  
DB 1 MSLVNRKQLEKMANVFRQEDYVAILDALBEYHNMSNTVVEKYKLKIDINSLTDIYI 60  
QY 61 NTYKSGRNKALKKFKEXLTMVELEKNNSLTPVEKNLHPITWIGQINDTAINYNOWKD 120  
DB 61 DTYKSGRNKALKKFKEXLTMVELEKNNLHPVEKNLHPVWIGQINDTAINYNOWKD 120  
QY 121 VNSDYTVKPYDSNAFLINTLTKTIIVESATNTLTESFRENLDNPEFDYNNKFRKMEIY 180  
DB 121 VNSDYNVNVFYDSNAFLINTLTKTVVESAINDTLTESFRENLDNPEFDYNNKFRKMEIY 180  
QY 181 DKQKHFDYKQIENPEFDIINIITKYLSEYKDLALNKYIEESLNKLTANNGNDI 240  
DB 181 DKQKNFINTYKQREENPELLIDDIIVKTYLSNEYSKEIDELNTYIEESLNKLTQNSGNDV 240  
QY 241 RNLEKFAEDLVRLYNOELVERNLAAASDIIRISMLKEDGGVYLDVILPGIQDLPKS 300  
DB 241 RNFEFPNGESFNLYEQELVERNLAAASDIIRISALKIGGYLDVDMPLGIQDLPFS 300  
QY 301 INKPSITNTSMEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSPESALSSKSKSEIF 360  
DB 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSPESVSLASKSKSEIF 360  
QY 361 LPDDIKVPLEKVIKAFANNSVINQALISLKDSYCDLVINQIKRYKILNDNLPSINE 420  
DB 361 SSLGDMESPLEYKIAFNKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPASE 420  
QY 421 GTDENFTMTKIFSDKLASISNEDNMFMKITYNLKVGPAFVRSTINLSGPGVYTGAYOD 480  
DB 421 DNDFTNTTTFDTSINAEANAGRMMELGKYLVRGFPDPVKTIIINLSGPAYAAAYOD 480  
QY 481 LLMFKDNSTNIHLLEPELNRFPPTKISQLTEQETISLWSFNQARAKSQFEYKKGYPE 540  
DB 481 LLMFKEGSMNIHLLEADLNRFISKTNI SOSTEQEMASLWSFDDARAKAQFEYKRNYPE 540  
QY 541 GAGEDNDLDFAGNTVDKQYVSKKILSSMKTRNKEYIHYIYVQLQDKISYASCNLFPSK 600  
DB 541 GSLGEDNDLDFSQNIYVDKEYLLEKISLARSERGIHYIYVQLQDKISYEAACNLFAK 600  
QY 601 DPVSSILYOKNTGSETAYVYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKGSEFNT 660  
DB 601 TPYDSVLFQKNIEDSIAYYYNPGDGEIQEIDYKIPSIISDRPKLTFIGHGKDEFNT 660  
QY 661 DTFANLDVDSLSEIETIILNKAADISPKYIEINLLGCNMFSYSIYAETYPFKLLKIK 720  
DB 661 DIFAGFDVDSLSEIAADLAKEDISPKSIEINLLGCNMFSYSINVEETYPFKLLKVK 720  
QY 721 DRUSELMPISQDSITVSANOQEVNRINEGKBEILDHSGKWNKBEIISKISSEYISF 780  
DB 721 DKISELMPISQDSITVSANOQEVNRINEGRRELLDHSGEWINKBEIISKISSEYISF 780  
QY 781 NPKENKIIVKSKYLHSLTLQEIIRNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840

DB 781 NPKENKIIVKSKYLPSTLQEIIRNANSSDIDLEKKVMLTECEINVISNIDTQIVEER 840  
QY 841 IEEAKNLTSDSINYIKNEFKLIESISDLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900  
DB 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLKQQLNELEDSHFISFEDISDEGFSIRF 900  
QY 901 INKETGNSIFETEKEIEFSEYATHISKEISINIKDIFDNVNGKLVKKNLDAHEVNTLN 960  
DB 901 INKETGESIFVETEKTIIESEYANHITIEISIKIGTIFDTVNGKLVKKNLDTTHEVNTLN 960  
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLPSTGLNTITDASKVVELVSTALDETD 1020  
DB 961 RAFFIQSLIEYNSKESLSNLSVAMKVQVYAQLPSTGLNTITDAKVELVSTALDETD 1020

RESULT 2  
US-10-011-366-6  
Sequence 6, Application US/10011366  
Publication No. US20030054493A1  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
Kink, John A.  
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
DISEASE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/011,366  
FILING DATE: 16-NO. US20030054493A1-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FILING DATE: 23-OCT-1997  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01121  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2710 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-011-366-6

Query Match 53.6%; Score 2783; DB 15; Length 2710;  
Best Local Similarity 52.1%; Pred. No. 3.2e-166;  
Matches 534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;



```
Db 399 TYCISNGISNTDDISRLDFTLVNKLSPVQDNDFKSTVEKNKRDI DLLKNTTISQKFGDRF 458
Qy 239 ---DIRNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOP 295
Db 459 QLRDINTLESFKKPDQFFYQQEMLLRWNYAAASDQVRINILKEGGYITDITDILPAYSD 518
Qy 296 DLPSINKPDSITNTSWEMIKLEAIMKY--EYIPG--YTSKNFMDLDEEVQSPSALS 351
Db 519 KVSQLINE--KSDDRFFEDKLRIISSESILSIKGEYSIKH--DGLDETTLNQNLNII-- 575
Qy 352 SKSDKSEIFLPLDDIKVSPLEVKIAFANNVI-----NQALISLSDSYC 395
Db 576 SEIEK---LTIDY--FKPVETKVRDTPKIFRYQKWNTENTWIRGNMFLMTHKSGKC 630
Qy 396 SDAVINQIKRYKILNDNLNPSINEGTDFTNTWKIFSDKLASISNE-----DNM 444
Db 631 IDFILSGQKQYLEL--QRINDNISYNNLFYTT-----EDLSLNVAIGGIPAKKYLEHG 684
Qy 445 MFMIKITNYLVKGFAPDVRSTINLSGP-----KDNLLTSSNDIV 471
Db 685 LF-----SEYRQDGTIYVVSNTINISGPDIMRQMKYKSLGRIGEVHDKNKLSDVNF 740
Qy 472 GVTYGAQDILMP---KDNSTNHLLEPE-----LRNFFPKTKISQJL-----EQEITS 518
Db 741 GVTASSNKONKSPWLNPNVSGINDITPDDESSWAVRNDINKILPEKINCHVPEKLP 800
Qy 519 LMSFNQARAKSQEEYKGYFEGALGEDDNLDPQNTVLDKDVYSKILSSMKTRKEYI 578
Db 801 LY-----YEIDSRSPFOGW---DNKSIKHVTEINKDIL---KDNLLTSSNDIV 844
Qy 579 HYIVQLQGDKISYEASCNLSKDPYSILYQKNIEGSETAYVYVADAIEKEIDKYRIPY 638
Db 845 KLUIKL--DRELVASKI-----DNPLALRSITLQLQLANVTSNTPENTINFY 896
Qy 639 QISNKRN-----IKLFTIGHGKSEFNTDTFANLVDLS-----LSSEIT 677
Db 897 DFVRKKQDILLSAUKL-----FSRNDADTKIIVWYNSVMEKNVFLREVISC 942
Qy 678 IMLAKADISPKYIEINLLCNMFSYSIYABEYYPGKLLKIKORVSELMPSISQSDI-- 735
Db 943 VLRSKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKMKELFSLDLDGYYK 990
Qy 736 TVSANOVEYRINEEGREILDHSGKWINKKEESTIKDISKEYISFNP-----K 783
Db 991 IITNAY---IKERDKL-----SGLIYNIENSIISGHESFDIIRSQHEWGDLSYEQPK 1042
Qy 784 ENKIIVKSKYHLSTLLOEIRNANSSDIDLEKKVNLTECEINVAS-----NIDQIYE 838
Db 1043 KPEFYVKSE--LSSAKSIFDDIKNK--YITDPTKRNVLXHQLDSDIKERIAFLDISHAYP 1100
Qy 839 GRTEBAKNLTSDSINYKNEFKLIESI-----SDSLYDLKHQNG 878
Db 1101 GSLEKLQLS---GYVFSIDINIABYLLASYGVSGHYGVVYPAPSPDKLLELRHTK 1156
Qy 879 DSHFISFIDISKTENGFRIRPINKETGNSI---PIETEKEIPSEYATHISKEISNIKT 935
Db 1157 SNSEW---EKIP-----YVVDILSDNVSNVLRPLSEEQKILNDIKLEISKVS---EQ 1207
Qy 936 IFDNVANGLYKYNLDAAEVNTLNSAFFIQSLIEYNTTK-----ESLSNLSV----- 983
Db 1208 YFMKLEQKSSVIGIKYVDFDRYNENLFLSLPQNLTLPFWRYVFEMLYDIHIGIEN 1267
Qy 984 -AMKVQYAOQLFSTGLNTITDASKVLEL 1010
Db 1268 KANREFIYKFSLSNLDPLINDERVNLN 1295
```

## RESULT 4

US-10-304-095-6

; Sequence 6, Application US/10304095

; Publication No. US20030134275A1

; GENERAL INFORMATION:

; APPLICANT: Long, David M.

```
; APPLICANT: Metz, Anneke M.
; APPLICANT: Love, Ruschelle A.
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
; FILE REFERENCE: 47714-5009-US
; CURRENT APPLICATION NUMBER: US/10/304,095
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: US/09/417,485
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2184
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (330)..(335)
; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-10-304-095-6
```

Query Match 5.0%; Score 261.5; DB 12; Length 2184;  
Best Local Similarity 19.7%; Pred. No. 1.3e-07;  
Matches 252; Conservative 193; Mismatches 442; Indels 395; Gaps 61;

```
Qy 10 QKVVYVKRIQDEYVAIILNALLEY--HNMS--SVVEKYLKLDI-----NLLTD 57
Db 572 KCIPIKLGSHNPKIFLKNVKKFLPNFKESFSLNQVMKNVKNIPQKKISKYNIKN 631
Qy 58 NYL--NTYKSGSRNKAL--KKFKEYLT---MEVELKNNSLTPVEKNLHFTWIGQIND 109
Db 632 RLLKNIPDNVYENKILHRNNKEIITINDNIKIYKNDNL-----NNSF-----KIKT 681
Qy 110 TAIN-----YINQWKVN-----SDYTVKVVYDSNAFLNTLTKTIVESATNNTL-- 154
Db 682 TLFNKLRRKYFNKIKKINAIQKRLMNLRIYFLF--NYFIMPLIRRRFFLTKEQTLHK 739
Qy 155 -----ESFR-----ENLNDPEEDY--NKF--YRGRMEIIVDKQH 185
Db 740 TIFPRKIWNHPTKISNFCLYHQIFRNKKLKRGNPFMDYQNMNVNKKKGKGI--KTNK 797
Qy 186 FIDYYSQIENPEFIIDNIITILSYNSYKDLAL---NKYIBESLNKITANN---GN 238
Db 798 YIFIKMKKSKSTNCKNNKFSKCKPKKKKKNLYNITRNNNIFIKDMKKSKTNLINK 857
Qy 239 DTRNLEKPAD--EDVRLYNQEL---VERWNLAAASDILRISMLKEDGGVYLDVILPGI 293
Db 858 SIDNLYKLKEINKSVRPYIKKYFYKIKKYPALKKMYIHMMAKEE-----904
Qy 294 QPDLFKSINKPDSITNTSWEMIKLEAIMKY-----KEYIPGYS-----KN 334
Db 905 -----KS-----NIKLERAFKFFIFAQEKELHULKYFSSHFQNRKKNYGR 946
Qy 335 FMDLDEEVQ-----RSFESALSSKSDKSEIFLPLDDIKVSPLEVKIAFAN 379
Db 947 FNKLHRIKNIILKQNSGVKNKDKTFLHLIKNKNKN-----N 986
Qy 380 NSVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINEGTDFTNTWKIFSDKLAS 439
Db 987 NKKKN-----KNYNN-----NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1031
Qy 440 NEDNMFMFKITNYLVKGFAPDVRSTINLSGPVVTGAYQDLL-----MFKDNSTNHL 494
Db 1032 NNNNNK---KAKNNEKNID---SNLEKKKKKIVYIKNIIEKRNFMKLKNSIN--HPI 1084
Qy 495 EPELNFEPFKTK-----ISQTEQEITSLWSFNQARAKSQFBEYKGYFEGALGEDD 547
Db 1085 SKKLRIINIPKKKGLRPLINLSTLVNPEIVKQRIEFLKSKSSE----FYHNLNLE 1140
Qy 548 NLDFQNTVLDKDYVSK-----KILSSMKTRN-----574
Db 1141 REKKDNIKKRYKNKQNFNPVSLNNICNFSKLCLGNMRHNNNSLFPNTLTKTGTIELKL 1200
```





QY 1002 TDASK 1006  
 Db 1351 TDYK 1355

RESULT 6  
 US-10-087-464-10  
 ; Sequence 10, Application US/10087464  
 ; Publication No. US20030059436A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chishti, Athar  
 ; APPLICANT: Oh, Steven  
 ; APPLICANT: Liu, David  
 ; APPLICANT: Goel, Vikas  
 ; APPLICANT: Li, Xuerong  
 ; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof  
 ; FILE REFERENCE: S1237/7019  
 ; CURRENT APPLICATION NUMBER: US/10/087,464  
 ; CURRENT FILING DATE: 2002-03-01  
 ; PRIOR APPLICATION NUMBER: US 06/272,930  
 ; PRIOR FILING DATE: 2001-03-02  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 10  
 ; LENGTH: 1639  
 ; TYPE: PRT  
 ; ORGANISM: Plasmodium falciparum  
 ; US-10-087-464-10

Query Match 4.8%; Score 249; DB 15; Length 1639;  
 Best Local Similarity 20.9%; Pred. No. 5.5e-07;  
 Matches 257; Conservative 215; Mismatches 418; Indels 338; Gaps 69;

QY 16 KRIQDEY-VAILN-ALREYHNMSESSVVEKVL-----KLKQINLITDNYLTYK----- 64  
 Db 275 KKQLYQAQDLSYNNKQLEAHLN--ISVLEKRIDTLKKNENIKELLDK-INEIKNPPEA 331

QY 65 KSG-----RNKAL-----KKFKEY-----LTMEVLEL-----KN----- 88  
 Db 332 NSGNTNTLLDKNKKTEEHEKEIKETAKYIKFNIDSLFTDPLELEYLREKKNIDISAK 391

QY 89 -----NSLT-PVEKNLHPFIWIGQINDTAINYINQKDVNS--DYT---VKF 129  
 Db 392 VETKESTEPNEYGVYPLSYN-----DINN-ALNELNSFGDLINPFDTKEPSKN 442

QY 130 VYDSN-AFLINTLTKTI-----VESATNTLESFRENLDNPEFFYKFKRMELIYD 181  
 Db 443 IYTDNERKKPFINEIKIEKKIES-DKKSIEDRSKSLN-----DITKEYEKLNEIYD 497

QY 182 -KQKHFDY--YKSOIEENPERFIIDNI--KTVLSNEYSK-DLEALN---KYIEE-SLAK 231  
 Db 498 SKFNNDIDLTNFKWMGKRYSYKVEKLTHTHTPASYENSKHLEKLTALKATKMTWEDYSLRN 557

QY 232 ITANN-----GNDIRNLEKFADELVLRYNQLVERWN-----LAAASDILRI 274  
 Db 558 IVEKELKYKYNLISKIENEIETLVENIKKQBEQLEKTKITKDNKPKDEKILEVSDIVKV 617

QY 275 SM-----LKEDGGVLDVILPGIQ--PDLFKSINKPDSINTNTWEMIKLEAIVK 322  
 Db 618 QVQKVLMMNKIDELKKTQLIKNLKVELKHNHVPNSYKQENKQEPY-----YLIVLKEIDK 673

QY 323 YKEIPIGYTSKNFMDLDEE---VORSFESALSSKSDKSEI-----PLPLDD 365  
 Db 674 LKVFMPKVS-----LINEEKNKIKTEGQSDNSPEFTEGKITQATTKPGQAGSALEGDS 729

QY 366 IKVS-----PLEVKI-----AFANNSVINQALISLSDKSYCSDLVINQIKN 405  
 Db 730 VQAQAQEQQAQPPVPVPEAKAQQVTPPAVNNKTEENVSKLDYLEKLYEFLNTSVYICH 789

QY 406 RYKILNDNLNPSINTEGDTFTNTWKIFSDKLASINEDNMFMKIKIINYLVKVGFPAPVRS 465  
 Db 790 KYILVS---HSTWNE--KILQYKITKEBESKLSKCDPLDLFLNQNIPVMYS--MFDS 842

QY 466 INLSQPGVYTGAYQ-----DLIMFKDNGSTNIHLEPELRNFEFPKTKIS-----QLTEQ 514  
 Db 843 LANSLSQLFMETIYKEMVCNLYKLDNDKIKNLLBEAKKVSIVTKLSSSSQPLSLTPQ 902

QY 515 EITSLSWSPNQAQAKSQPEYKGYPEGALGEDDNDLDFPAQNTVLDK--DYVSKILSSMKT 572  
 Db 903 DKPEV-SANDDTSHSTNLNLSKLFPENILSLGKNKIYQELIGQSSNFYEKILKDSOT 961

QY 573 RNKEYIHVQLQGDKISVEASCNLFSDPYSSILYQKNIEGSETAYYYYVADAEIKED 632  
 Db 962 FYNESFTNFVSKADDIN---SLNDESK-----RKKLE-----EDIN 995

QY 633 KYRIPIQIS---NKRNIKLTFIGHGKSEFNTDTFANLDVDSLSSEIETILMLAKADISP 688  
 Db 996 KLKKTQLQSFDLYNLYKLERLFDKKTGVGYKMQIKKLTLLKQESKLSLN---NP 1052

QY 689 KYIEINLGCNMFSYIY-----ABEYTPG-KLLK-----TKORVSELMPP--S 729  
 Db 1053 KHV-----LQNFSPFNKKKEAIEAETENTLENTKILLKHYKGLVKYNGESSPLKT 1104

QY 730 ISQDSITVSAN-----QYEVRIINEBK-----REILDHSGKWINKESIIKD-ISSK 775  
 Db 1105 LSEESIQTEDNYASLENFKVLKLEGKLDNLEKKKLSYLSGHLHLIAELKEVIRNK 1164

QY 776 EYISFNPKNKIIVKS-----KYLHE---LSTLQEIERNNANSSDIDLEKKVMLTEC-- 824  
 Db 1165 NYTGNSPSENNTDNNALSYKFLPEGTDVATVSE--SGSDTLEQSQPKKPASTHVCA 1222

QY 825 ---EINVASNDRQI-----VEGRIBEAKNLTSDSINYIKNEPKLIESISDSLYDLKH 874  
 Db 1223 ESNTITTSQNVDDVDVIVPIFGESE--DYDDLQGVVVTGEAVTPSVIDNI----- 1273

QY 875 QNGLDSDHFISPEDISKTEGPNRIFINKETGNSIFIEETEKEIFSEYATHISKEISNID 934  
 Db 1274 -----LSEIENEYEVLYLKLPLAG--VYRSLKQLENNVMTF-----NVNVKD 1313

QY 935 TI-----FDNV-NGKLVKVNLDAAHEV-----NTLSAFFIQSLIEYNTTKESL- 978  
 Db 1314 ILNSRFNKNRENFKNVLESDDLIPYKDLTSSNYVVKDPYKFLNKEKDKFLSSYNYIKDSID 1373

QY 979 SNLSVAMQVYQAQLFSTGLNTITDASK 1006  
 Db 1374 TDINFANDVLGYKILSEKYSKSDLSIK 1401

RESULT 7  
 US-10-032-585-7646  
 ; Sequence 7646, Application US/10032585  
 ; Publication No. US20030180953A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Terry, Roemer D.  
 ; APPLICANT: Bo, Jiang  
 ; APPLICANT: Charles, Boone  
 ; APPLICANT: Howard, Bussey  
 ; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
 ; FILE REFERENCE: 10182-005-999  
 ; CURRENT APPLICATION NUMBER: US/10/032,585  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 8000  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 7646  
 ; LENGTH: 1881  
 ; TYPE: PRT  
 ; ORGANISM: Candida albicans  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (1881)..(1881)  
 ; OTHER INFORMATION: X-any amino acid  
 ; US-10-032-585-7646

Query Match 4.7%; Score 246; DB 12; Length 1881;  
 Best Local Similarity 19.9%; Pred. No. 1e-06;  
 Matches 234; Conservative 187; Mismatches 360; Indels 396; Gaps 56;

Qy	13	VYVFRIODEYVAILNALEYHN--MSSESVVEKYLKJOIN-----NLTNDYLNATYKKS--66
Db	697	VYV--NLVKDNFLKIKRAL--FHPDSSPRGKISYELFEELDTKVLNLTRE--LQTERENA751
Qy	67	-GRNKALKFKXEYLT-----MEV--LELKNLSLTPVEKNLHFIIWIGGOINDTAINYQWK119
Db	752	ESNDKELNEKTEKLTNLSTKLETKLEDKEQELAKIQED-----HK791
Qy	120	DVNSDYTVKFVYDGNAF-----LINTLKKTIVESATNNLTLESFRE159
Db	792	SLNE-----KFLVTANSLOGIKARTKESETISGPDQOEQLKAKK--GNTSESTLKQKE844
Qy	160	NLNDPEFFYNK-----FYRKMEI-----YDKQKHP--IDYKSOQIEENPEF200
Db	845	KLDSTEQAQKLEDGINNMTRDLFHLKKSKEAEATQIKQREREFKNLTYPENTKKOYEL904
Qy	201	LIDNIIKTYLSNYSKOLEALNKYIEBSLNKITANNQNDIRNL--EKPADEDLVLYNOEL259
Db	905	QINNUNKS--NNEFQKINELUSKKIE-----SLTEBNKFNKQLEKURDTE---ENNEHL955
Qy	260	VERWNLAAASDILRAISMLKEDGVYLDVILPGIQPOLFKSINKPDSITNTSWEMIKLEA319
Db	956	MDKLRSASVA-----YNDLKKAKSESE-----EE979
Qy	320	IMKYKEYIPGVTYKSNFMDLDEEV-----QRSFSAJSSKSDK--SEIFLPLDD---IKV368
Db	980	TVKAKEELETLSK-IDNLEKELKEQQSKNKELEGQLQNTIDSTNEKFKLEDELEKSIKK1038
Qy	369	SPLEVKIAPANNVINOALISLKDYSCLDVINOIKRYKILNDNLNAPSINEGTDFTWYM428
Db	1039	SNKE--ISSQNSLETQLEKTEKOLQAKDEBIDKLKAETKSNIDNLNASEI---SSLOSKL1093
Qy	429	KIFSCLKIASIGNEDMMFMKIITNVLKVGFPAPDVRSTINLSGPGVYTGAYDQLLMPKONS488
Db	1094	KEAESHSSTKDEHSSL-----S1111
Qy	489	TNIHLLEPELNFBFPKTKISQLTE-----QEITSLWSFNQARAKS529
Db	1112	ENLKKLKEEYEN-----TKTSWIAKLSAKIBEHKKATDEIETKTKHITDLQE--SHAKQKS1165
Qy	530	QPEEYKGYFGALGEDNDLPDAQNTVLDKDYVSKKILSSMKTRNKYIHYIVOLQGDKI589
Db	1166	QFESERNDI-----KSNLDEANKELSD-----NREKLSNLE--REKTELANKLQTOBEKI1213
Qy	590	SYEASCNLFKDPYSSILYQKNIEGSEYAYYYYVADAEIKEDIKVRYPIQISNKRNTKLT649
Db	1214	SDLETSVAISDCKSKLKH-----DIEOLKKEKIKLETTLKENSETM1255
Qy	650	FIGHKSFEYNTDTPANLD-----VDSLSSIEFTI--LNLAKADISPKEIENLLGCNM700
Db	1256	FEKKEQLQVNDKCKELEACLKLTETKEKENDLIRKLEAAKSD-----1300
Qy	701	FSYSIYABETYPGKLLKIKORVELMPSISQDSITVSANQYEVRIINEGKREILDHSGK760
Db	1301	-----HDTERRKKLSLLIEDTKSE-----SEKNVINKLNEQIEKLXGEREKEVRDIQSQ1347
Qy	761	-----W-----INKEESIIKDISKEYISFNPKENKIIVK-----S791
Db	1348	LAAKTTDMEKIKTTLDKVLKESKSULEK--TNKESVDTLKKEVENLKGEISLLEDOKKDDTT1406
Qy	792	KYLHELSTLLQEIERNNANSDDILEKKVMLTECEINVASN-----IDROIVE838
Db	1407	KY--KELAAQLETKTSNLDSTTMELEK-----TELELKKVRNELTEATSELTQLQNNQSILT1461
Qy	839	GRIBEA-----KNLTSDSINYINKNEFKLIES--ISDLSYDLKHXQNLGDD880
Db	1462	BEIEKTKAALTSSKOLEVCGNQSKSLQDSLVKSELKNPENKYNQETTSKLDEIBEKQ1521
Qy	881	SHFTISPDISKTENGFRIRPFINKETGNSIFITEKEIFSEYATHI-----S926
Db	1522	KEIYTLQ-----TELKDRISEYEKE--RAMLSENSETVIKEYSDIKSLGKINSIKENHS1575

```

QY   927 KEI-----SNIKDTI-----FDVNGKLVKKVN-----LDAAEHVNTLNLSAFFTQSLLI  969
    |||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db   1576 KEITHNEQTSLKQDIAKLSQDHESAQTQLDKENQKLKASLEKHNTESA---TSIE 1632
    |||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
QY   970 EYNTTKESLNSLVAMKVQVYAQLFSTGLNTITDASK 1006
    |||      :|||      :|||      :|||      :|||      :|||      :|||      :|||
Db   1633 EKNNQIKELSETIKSLK-----TELKTSGDALX 1660
    |||      :|||      :|||      :|||      :|||      :|||      :|||      :|||

RESULT 8
US-10-032-585-7127
; Sequence 7127, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7127
; LENGTH: 1270
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (340)..(340)
; OTHER INFORMATION: X=any amino acid
US-10-032-585-7127

Query Match          4.3%; Score 224; DB 12; Length 1270;
Best Local Similarity 17.7%; Pred. No. 1.4e-05;
Matches 220; Conservative 207; Mismatches 399; Indels 416; Gaps 51;

QY   15 VKPRIQDEVAIVNALNEEYHNMS-----ESSVVVEKYKLKDINNLTNDNYLNTY---KKS  66
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   42 IKYYTOE-AKQLTTAIE---NLSMVNEESDIKKKYFLVNIINLRQDPFKYVTLIKWA  97
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   67 GRNKALKPKXYL-----TNVEL-----84
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   98 SISKDVSKFIDLLNWFIQBFHPFENLIFQLNALTGYNSAKLPNSDIITALEVLHYHGRPKL 157
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   85 ----ELKNLSLTPVKNLHIWIGGQINDTAINYINQWKDVNS-----DYTVK----FVY 131
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   158 PSVNYIKSDNLSP-QKILE-----TLNDLNLVMTFRFALMDNI PKRFDEYEIKDGRAYIR 210
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   132 DSNAFLN-----TL 141
    |||  :|||
Db   211 VSNFEFVSITVGNDLIIDNPVEYKSPFFYDFDKFLFGANPESGLTFNDDDKISTKLP TS 270
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   142 KKTIVESATNNLT-----ESFRENLNDPEFD---YNKFYRKMEIYYDK 182
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   271 SHKKLEKLVNQTLTRGLQGLYELLHKYSNFXYLLAQOFOTLLINSRWGNFNQYNYQT 330
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   183 QKHFDY-YKSO--IEENPFIDNIKIYLSNEY-----SKOLEAL-----221
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   331 NKSLIVINYXSQHLYSLRNWNNSFIELGDITHSNLNRYWFKNQGYCFDQGQNNELDKIFHLQ 390
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   222 ---NKYIBESLANKITA-----NNGNDIRNLEKFADEDL--251
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   391 RRNSNGVTTSSNTISTISAMIAGRSTELSNETAIVDDNDNDNDNTNTGETENEDLN 450
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   252 -----VRLYNQBLVERWNLAASAADI-----LRISMLKED-----GG 282
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   451 DLILNVVNVQHAKSINSEIYSQ-LLFTEFSETDYSWVSPHQLLLOISPCKSTVFAINPLTG 509
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
QY   283 VYLDVDILPGIQDPLFKSINKPSOITSNTWSMEMIKLEAIMKYKBYIPGYTSKNFMDLDEEV 342
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db   510 FFVIDPTP-IQTYITKKINSPPPT-----LSOVSIKOSFIP-----ESDMISYVI 555
    :|:  |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

```





```
; SEQ ID NO 419
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-419

Query Match      4.1%; Score 213; DB 15; Length 1979;
Best Local Similarity 19.6%; Pred. No. 0.00013;
Matches 227; Conservative 197; Mismatches 417; Indels 318; Gaps 57;

QY 4 VNKAQIQKMYVVK-FRIQDEYVAILNALAEYHNMSSESVVEKYLKLD-INNLTDNYLN 61
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 VHQLEDDKMDITKELDVQKEKLIQSEVALNDLHLTKQ-----KLEDKVENLVQD-LN 593
QY 62 TYKSGRNKALKFKFYLFWELKGNLSLTPVEKNLHFIWIGGQINDTAINYINQWQDV 121
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
594 KSOES--NVSIOK-----ENLEKEH-IRQNEEL-----SRINELMQSLNQ--DS 635
QY 122 NSDYTVKVVDSNAFLINTLKKTIVESATNTLTSPRENLDNPEFDYNYFKRMEIYD 181
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
636 NSNFKDTLKERAEVRN-LKQNLSE-----LEQNLNMLKKVAFDV-KWNEKLVACE 687
QY 182 KQKHFIIDY---KSQIEENPEFIIDNI-----IKTYSNEXSKOLEALNKYIEESLNKI 232
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
688 DVHQLEECCLAGNNQLSLEKQNTIVETLKMEKGEIAELCWAKKRLLEEANKY-EKTIEL 746
QY 233 TANNGNDIRNLEKFA---DEDLVRLYNQELVERNWL-----AAASDILRISML 277
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
747 -----SNARNLNTSALQLEHEHLIKLNQKQDMIEAELKKNIEQMDTDHRETQDVLS-SSL 800
QY 278 KEDGGVYLDVDPGIGQDPLFKSINKPDSITNTSWEMIKLEALMKYKEVPGYTS---KN 334
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
801 EE-----QKQLTQILNKKEIFI-----EKLK-ERSSKLOEBELDKYSQALRN 841
QY 335 FDMLEBEVQRPESALSSKSDSEI-----FLPLDDIKVSPLEVKIAPFANS 381
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
842 -EILRQOTIEKRSLSGMBEENNHQBELERLEEQRSTAPVADPKTLDSTVELA---S 896
QY 382 VINOALISLSDSCVSLVINQIKRNYKILND-----NLNPSINEGTFD----- 425
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
897 EVSQ-LNTIKEH-----LEEIKHKKIIEBQNSKXQMLLSQLOEQKEMDEFYQHEQM 950
QY 426 --FTWKIFSDKLASISNEDNMFMKITYNLKVGFPAPDVRSTINLGGPVYTGAYQDLWM 483
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
951 NATHQOLFLEKDEIKSLQKTIEQIK-----TQLHEERQD 985
QY 484 FKNSTNIHLLEPELRNFEPPKTKISQLTEQETISLWSPNQARAKSQFEEYKGYFEGAL 543
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
986 IQTDNSDI-----FOETKVQSLNIENGSEKHDLSKASTERLVKGIKERELEIKL 1034
QY 544 GEDDNLDFQA-QNTVLDKDYVK-----KILSSMKTRNKYIHYIYVQLQ 585
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1035 LNEKNISLTKQIDQLSKDEVGKLTQIIQQKDLBIOALHARISSTSHQDVVYVLLQ--OLQ 1092
QY 586 GDKISYEASCNLFSKDPYSILQKNIEGS--ETAYVVYVADAEIKEDKYRIPYQISNK 643
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1093 -----AYAMEREKVFALWNEKTRENSHLTEYHKMDIYVAKEALIKL--QDENK 1141
QY 644 RNIKLTFIGHKSEFNFTDPANL-----DVDLSSEIETILNLAKAD-----I 686
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1142 K-LSTRFESSQDMFR-ETIQNLRIIREKDIIDALSQKQTLAVLQTSSTGTNEAGGV 1199
QY 687 SPKYIEINLLGCNMFYSIYABETYPGKLL-----LTKIDRVSELMPSISODS-- 734
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1200 NSHQFELLOERDKLKQOVKMEEWKQOVMTTVQNMQHESAQLELHQLQAQVLVDSN 1259
QY 735 -----ITVSANQYVRINEEGKR-----EILDHSGKWNKEESIIKOISSK 775
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1260 NSKLQVDYTGILQSQYEQNEFKLNFKGQELAQVQHSIGQLCNTKDLGLKLDLISPQLSSA 1319
QY 776 EYISFNPKNKIIVKSKYLHELSTLQ-----EIRNNANSSDIDL-----EKKVMLTECEIN 827
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1320 SLITPQSAECLRAKSEVLSESELLOQEELEELKRSLOEKDATIRTLQENNHRLSD-SIA 1378
pb : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 828 VASNIDROIIVEGRIBEAKNLTSDS---INVIKNEFKLIESISDSLYDLKHQGLDSDSHFI 884
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1379 ATSELERKEHQTDSEIKQLKEKQDVLOKLKKEKDLLLAKKSDQL----- 1423
QY 885 SPEDISKTENGFRIRFINKETGNSIFITEKEIFSEYATHISKEISNIKDTI--FDNVNG 942
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1424 ---LSSNEN-----FTNK-----VNEELLRQAVTNLKERILILEMDIGLKGGENE 1466
QY 943 KLV-----KKNVLDRAHEVNTLNSAFFIQSLIEYNTTYESLNSLVAMKVQVYALF--- 994
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1467 KIVETYRGKETEYQALQETNMKFSMMLREKFECHSMKE-----KALAFEQLLKBEK 1517
QY 995 ---STG-LNTITDASKVVE 1009
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1518 EOGKTGELNQLLNAVKSQ 1536
US-10-155-533-9
RESULT 12
US-10-155-533-9
; Sequence 9, Application US/10155533
; Publication No. US20030104003A1
; GENERAL INFORMATION:
; APPLICANT: Nguyen, Thanh V.
; APPLICANT: James, Anthony A.
; TITLE OF INVENTION: A No. US20030104003A1el Surface Protein of the Malaria Parasite
; FILE REFERENCE: 48417/CAB/R2682
; CURRENT APPLICATION NUMBER: US/10/155,533
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1610
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-155-533-9
```

```
Query Match      4.1%; Score 212; DB 15; Length 1610;
Best Local Similarity 18.6%; Pred. No. 0.00011;
Matches 249; Conservative 200; Mismatches 435; Indels 454; Gaps 64;

QY 11 KMVYKFRQDEYVAILNALAEYHNMSSESVVEKY-----LKLKLDIN----- 53
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
339 QMSYNNINTNSNEYDKNASTLDQTY---IGKTFEGVYVSVNKAACIKLNINKYGLLPK 395
QY 54 ---NLTDNY-----LNTYKSGRNKALKKFKFYLFWELKGNLS 91
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
396 NKANLGDDIEDMNDPFKQDPVHVVKILGINTKKNIFYLGNIIKYANENIKLSKGEYSKGLI 455
QY 92 TPVEKNLHFIWIGGQINDTAINYINQWQVDSYTVKVVDSNAFLINTLTKKTIVESATN 151
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 TKVCSYCFI-----KVLKNGSTGYLHKSLFCMNDKEK-----KN 491
QY 152 NTLSEF-----RENLDNPEFDYNYFKRMEIYDQKH---FIDYYS 192
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
492 NDQNVDPNPNPNVQNVQNVNPNYDHPYDQNVQNVNPNPNYDQNVQNVNPNPNYDQNV 547
QY 193 QTEENPEFIIDNIITYLSNEYS---KDEAL-----NKYIEESLN 230
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
548 QLYNSDNLMQDPIYQLQFTKIFNIWDIIDVEILGTQPDNDYKNGNYLTI PRGSKTFKILN 607
QY 231 KITANNGN--DIRNLEKFADEDLVRLY-----NOELVERWNLAASDI--LRISMLKEDG 281
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
608 YUNVLKENEIDNNIQKGD-----YYISDNNQNDNIID-----SDINHHNNKKKKK 656
QY 282 GYVLVDVILPGIQPOLFKSINKPDSI-----TNTSWEMIKLEAIMKYEYIPGYTS 332
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
657 NLY---DIQNNMNSHPFNKFHTEDEYLFNDHVQENVHTEFE-----KNKKYKITDYK 705
QY 333 KNFDMLEDER--VORSPESALSCKSK--SEIFLPLDDIKVSPLEVKIAPFANSVINQALI 388
Db  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
706 ENNHMKNSYLLKQNKELPFNNKFKKIIKNIYDLPNTISLSMLSKTIKIPLASIKKYFII 765
```

QY 389 SLKDSYCSDLVIN--QIK-----NRYKI-----LNDNLNPSINEGDTFNTMTKIF-----SD 433  
Db HENKEYNSSYKINSEIKRICQFKIDCNVEQDQNVVTKVN--GTTKDCQEKVFNVTQD 824  
QY 434 KL-----ASINEDNMFMKLT---NYLKVGFPDVRSTINLSGPGVYTAY 478  
Db RUKEGEQERVIKVEAKIKNDWMVQOKDKKEKMDVQFIEE--KDIINVQHINQDMVDV 882  
QY 479 QDLMPKDNSTNHLLELRLNFPFKTKISOLTEQETSLWSFNQARAKSQFEEYKGY 538  
Db 883 QDM-----DVQINVQDMVQINQDINQDMVQINNSITLWKS--TSCQDSESR--- 933  
QY 539 FPGALGEDNDLQFAQNTVLQDKYVSKILSSMKTRNKYIHYVQIQGDKISYEASCNLF 598  
Db 934 --DAPGGQVESLDEKDSMEKSEKKEKKKGRKKNKD--TNLTLSKDSIQ--KSKTTLD 987  
QY 599 SKDPYSSILYQKNIESEFAYYYV--ADAEIKE---IDKYRIPYQISNKRNIKLTFI-- 651  
Db 988 DKGRNVVVTFIGHINHGKTSLFDYICKTNEQKKEYGLITQNTIRAFKATVRNNFTFLVDT 1047  
QY 652 -GHGK-----SEFNTD-----TFANL 666  
Db 1048 PGHEAPMPMRSGVKISDLSILVISGDEGIQGTVECIKLKEFNKIIIAITKVDPNV 1107  
QY 667 DVDSL-----SSIEITI-----LNL----- 681  
Db 1108 DVDIRIINDLYHDITTELNGBEIQVVECSIYKEESIDKLDAIYLESEFLNLQTNPKKH 1167  
QY 682 --AKADISPKYIE-----INLL--CCNMFSYIAEETYPGKLLKIKIDRVSELMPS-- 729  
Db 1168 EQAQGVVLSYIDKNGIVSINLLQNGVLNHDHFYTGSSY-GKVKT-LKDHKNKNIKISAY 1225  
QY 730 -----ISODSITVSANOVEVRINEEGKEIILDHSGKWINKEBSIIKDISSEYISF 780  
Db 1226 PSDPIKLIQYKNKNSVPAGDKFVYVENEALAEIAEH-----NKNKMLTWEINNFTYDQ 1280  
QY 781 N-----PKENKIIVSKYLHE----- 796  
Db 1281 NMNRYKDFIISRENKIGSGSILGENNLKNDIDGNMTRDDNMTRDDNMTRDDNMTRDGNR 1340  
QY 797 -----LSTLQETR-----NNANSSDIDLEKKWMLTBC 824  
Db 1341 TNNDNITSDNNSNDYDKIKETKMYTNNKSFOKDDFLKHLNNTNENVMPPSTHIGKN 1400  
QY 825 EI-----NVASNIDROIVEGRIBEAQNL-----TSDSINVKNE--FKLIESISDLYDL 872  
Db 1401 EIKTIYSNVIKCDKO--GSIEVLKNCMLKLOKEDSICKKNKIYADIIGNWTSS--DI 1455  
QY 873 KHONGLDDSHPISF-----EDI-----SKTENGFRIRFINKETGNSPIFETEKEIPS 919  
Db 1456 KYATSP-NATIIAFVKLSNDIKGSKNSKGNHNNYPIYSN-----VLVELIENVEK 1508  
QY 920 EVATHISKE-IGNIKDT-----IPD-----NVNGKLKVKKNLDAAHEVNTL--NSAFFTQ 966  
Db 1509 ENEKLSKXPMGELKGTQAILKVPNISKLGVAGCIVKRGTTISINSNIRILNDRKVIYMG 1568  
QY 967 SLIEYNTTKESLSNLSVA 984  
Db 1569 KIISIKIVAEKQTQVTEA 1586

RESULT 13  
US-09-815-242-5809  
; Sequence 5809, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5809  
; LENGTH: 993  
; TYPE: PRK  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5809  
Query Match 4.0%; Score 209.5; DB 9; Length 993;  
Best Local Similarity 19.2%; Pred. No. 8.4e-05;  
Matches 206; Conservative 159; Mismatches 354; Indels 353; Gaps 56;  
QY 9 LQKVVVVFRIQDEY---VAIINALBEYHN-----MSESSVVEKYLKLDINNLTDN 58  
Db 115 IQK--YVKV---DGEWLKLVSYIESIDEYYQNKVNSKAHIINDRIYL--NEQSAI 166  
QY 59 YLNTYKSGRNKALKKFKKEYLTMEVLELNKNSLTPEVKNLHFIWIGQINDTAINYINQM 118  
Db 167 YLNNNKOTS-----FSIKNSL-----LVFKITVTNNITFSNLA 202  
QY 119 KDVSNDYTVKVVYDSNAFLINTLTKTIVESATNNTLESFRENLDNPEFYNK-----F 171  
Db 203 EKINGEFDINDITKVKVYIHNLVSKSEIYITRPP-L-SYSDNLN---YLNLKLSLHNDDF 258  
QY 172 YKRMETIIVDKHFDIVYKSOIBENPEFIQNI-----IKTVLSNEYSKDLEALNK 223  
Db 259 VKKIREI---QKLILAVEKTEIGFGEELYKDIHHMKALFKCKNYL--QIDTKIDMNN 312  
QY 224 YIEESLNKITANNNGNDIRNLEKFADESLVRLYNQBELVERWNLAASDILRISMLKEDGV 283  
Db 313 YLHQDI-----ATNISEA-----YLLWLSRNNIG----- 338  
QY 284 YLDVDILPGIQDLPKSIKNSDPSINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEE--- 341  
Db 339 FTDLKVL-----HNRFEIKYGFQELVNIKDLISDITGFGTTIQQEET 381  
QY 342 -----VORSPESALSSKSDKSEIFLPLDDIKVSPLEVKIAFANNVSNQALISKDSY 394  
Db 382 DGNNTVMLKQFLHALRNN-----DEIVINEKQVE--SLINDNEINHYPMSADV 430  
QY 395 CSDLVINQIKRYK--ILNDNLNPSINEGDTFNTMTKIF--SDKLASISNE----- 441  
Db 431 YAEVLGRFYQYNELIVISPLTASFNAGATGRFHHLLIDTETLAKLEHKGHYQKMIC 490  
QY 442 DNMFMKITYN-----LKVGFPDVRSTINLSGPGVYTGA----- 477  
Db 491 DDNEMISINNIIPKYPNRNHNLTNHDSEYSLNLG--SSNSYSKYELTLLDDIYVGATFNKL 549  
QY 478 --YQDLL---MFKDNS-----TNHLLLEPELBNFEPKTK--ISQLEQETISL--- 519  
Db 550 YLYSSQLNKRVLFSNNMNYNFKCNLYRL---LREISMESVKCIEPMNDVSDISFSYSP 606  
QY 520 -----WSFNQ-----ARAKSQFEEYKGYFEGALGEDDNDLQFAQNTVLD 558

Db 607 RIRYKXVILKPAWKINEMVLPKNEWDQFLKYQ-----EQFNIPNIVNLV-- 655  
QY 559 KDYVSKILSSMKTRNKEYI-----HYIVLOQDKISYASCNLFSDOPY 603  
Db 656 --YGDNKLNLNLNLANHRYLLMKEYKKHVRVLVESFLPQSKNDHV-YEIVTPIYKSSY 712  
QY 604 SSILYQKNIEGSETAYYYVADAIEIKIDYRIPYQISNKRNIKLTFIGHGSEFNTDTF 663  
Db 713 -----RGPEI-----EIPKY-----NTDIEYDKDWF 734  
QY 664 A-NLDVDSLSEIETILNLAkadISPKYIEINLGCNMFYSIYAEETPGKLLKIKDR 722  
Db 735 ALHIHIDKPSQDTFIIDNL-----YFVKHLKDKGD 765  
QY 723 VSE--LMPISQDSITVSANQYEVRI--NEEGKREILDHSGKWINKESIIKDISSKEYI 778  
Db 766 IDQYFLMRYIKQGD-----LKLRLFRNDENVAEYISILKNWLPVHRQ-TTEVSDFEYF 818  
QY 779 SFNPKENKIIVKSYLHELSTLLQEIEN--NANSSDIDLEKKVMLTEC-----EINVAS 830  
Db 819 SYEPFPRYGGKNT-INETEAFFEDYTNLAVNIENDFKFDRPYIVAISIMYLFEMFIS 877  
QY 831 NIDR-QIVEGRIEAKNLTSDSINYIKNEFKLTIESDSIL-YDLKHQNGL-----DDSHF 883  
Db 878 NEERMEIVNNYV-PTSFKSKDIRPFPKNELVTICNANNFEYMAKHSYGIYRILKDGNOI 935  
QY 884 IS--FEDISKTEGFRIRFINK---ETGNSIF-IETEKEIFSEYATHISKEI 929  
Db 936 LSKLNEGLKLTITKRSRIIGSLIHMCRNRIFGIDKQDTF---VLSIVKEI 984

## RESULT 14

US-09-815-242-12899  
; Sequence 12899, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIORITY APPLICATION NUMBER: 60/191,078  
; PRIORITY FILING DATE: 2000-03-21  
; PRIORITY APPLICATION NUMBER: 60/206,848  
; PRIORITY FILING DATE: 2000-05-23  
; PRIORITY APPLICATION NUMBER: 60/207,727  
; PRIORITY FILING DATE: 2000-05-26  
; PRIORITY APPLICATION NUMBER: 60/242,578  
; PRIORITY FILING DATE: 2000-10-23  
; PRIORITY APPLICATION NUMBER: 60/253,625  
; PRIORITY FILING DATE: 2000-11-27  
; PRIORITY APPLICATION NUMBER: 60/257,931  
; PRIORITY FILING DATE: 2000-12-22  
; PRIORITY APPLICATION NUMBER: 60/269,308  
; PRIORITY FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12899  
; LENGTH: 1002  
; TYPE: PR1  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12899

, Query Match 4.0%; Score 209.5; DB 9; Length 1002;

Best Local Similarity 19.2%; Pred. No. 8.5e-05;  
Matches 206; Conservative 159; Mismatches 354; Indels 353; Gaps 56;  
QY 9 LQKMYVVKRIQEDY-----VAILNALBEYHN-----MSSESVVEKYLKLDINLITDN 58  
Db 120 IQK--YVKV--DGEWLKLVSYIESIDBYQNLKVIWNSKAHIINDRIYL---NEQSAI 171  
QY 59 YLNTYKSGRNKALKPKPEYLTMEVLELKNLSLTPEVKNLHFEIWIIGGQINDTAINYINOW 118  
Db 172 YLNNNKOTS-----FSIKNSL-----LVFIKTTVTNNNITFSNLA 207  
QY 119 KDVSNDYTKVFDYDQNAFLINTLTKTIVESATNNTLESFRENLDNPEFDYNK-----F 171  
Db 208 EKINGEPEINDITKVKVIHNLVSKIEIYSTIRPPL-SYSDNLN---YLNLKLSLHNDFF 263  
QY 172 YKRMETIYDQKHFDYKYSQIEENPEFIIDNI-----IKTYSNEYSKDLALNK 223  
Db 264 VKKIREI---QKLILAVEKTEIGFGEELYKDIHHMKALPKCKNYL--QIDTKIDMINN 317  
QY 224 YIEESLKNITANGNDIRNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGV 283  
Db 318 YLHQDI-----ATNISEA-----YLLMLLSRNNIG----- 343  
QY 284 YLDVDILPGIQDLPKSIKPKDSITNTSWEMIKLEAMKYKEYIPGYTSKNFMDLDEE-- 341  
Db 344 FTDLKVL-----HNRFEIKYGFELQVNIKDLISDITGFGTTIQQEET 386  
QY 342 -----VORSPEALSKSKSEIFLPDDIKVSPLEVKIAPANNVINQALISLKDSY 394  
Db 387 DGNNIWMLKQKFLHALRN-----DEIVINEKQVE-SLINDNEINHYPMSADV 435  
QY 395 CSDIIVNIQKNRYK--ILNDNLNPSINEGTDNTMTKIF-SOKLASISNE----- 441  
Db 436 YAEYLGRPYNQYNELIVISPLTASFNAGATGRFHHLIDTETLAKLEHKGHYQKMTIC 495  
QY 442 DNMMFMKITNY-----LKVGFAPDVRTINLSGGVVTGA----- 477  
Db 496 DDNVEIMISINIPKYPRHNVLTNHDSEYVSLNLG-SSNSYSKYELTLDLDIYVGATFNKL 554  
QY 478 --YQDILL--MFKONS-----TNHILLEPELRNFEPKTK-ISQLTEQBITSL----- 519  
Db 555 YLYSQLNKRVLFESNNMNFLECNLYRL---LREISMESVKCIEPMNDVSDISFSP 611  
QY 520 -----WSFNQ-----ARAKSOFEPYKGYGEGALGEDDNLQFAQNTVLD 558  
Db 612 RIRYKXVILKPAWKINEMVLPKNEWDQFLKYQ-----EQFNIPNIVNLV-- 660  
QY 559 KDYVSKILSSMKTRNKEYI-----HYIVLOQDKISYASCNLFSDOPY 603  
Db 661 --YGDNKLNLNLNLANHRYLLMKEYKKHVRVLVESFLPQSKNDHV-YEIVTPIYKSSY 717  
QY 604 SSILYQKNIEGSETAYYYVADAIEIKIDYRIPYQISNKRNIKLTFIGHGSEFNTDTF 663  
Db 718 -----RGPEI-----EIPKY-----NTDIEYDKDWF 739  
QY 664 A-NLDVDSLSEIETILNLAkadISPKYIEINLGCNMFYSIYAEETPGKLLKIKDR 722  
Db 740 ALHIHIDKPSQDTFIIDNL-----YFVKHLKDKGD 770  
QY 723 VSE--LMPISQDSITVSANQYEVRI--NEEGKREILDHSGKWINKESIIKDISSKEYI 778  
Db 771 IDQYFLMRYIKQGD-----LKLRLFRNDENVAEYISILKNWLPVHRQ-TTEVSDFEYF 823  
QY 779 SFNPKENKIIVKSYLHELSTLLQEIEN--NANSSDIDLEKKVMLTEC-----EINVAS 830  
Db 824 SYEPFPRYGGKNT-INETEAFFEDYTNLAVNIENDFKFDRPYIVAISIMYLFEMFIS 882  
QY 831 NIDR-QIVEGRIEAKNLTSDSINYIKNEFKLTIESDSIL-YDLKHQNGL-----DDSHF 883  
Db 883 NEERMEIVNNYV-PTSFKSKDIRPFPKNELVTICNANNFEYMAKHSYGIYRILKDGNOI 940  
QY 884 IS--FEDISKTEGFRIRFINK---ETGNSIF-IETEKEIFSEYATHISKEI 929



Db 941 LSKLNEGLKTLTTKRSRIIGSLIHMRCNRIFGIDKQDQTF---VLSIVKEI 989

RESULT 15

US-09-815-242-13158

Sequence 13158, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zykkind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 13158

LENGTH: 1002

TYPE: PR

ORGANISM: Staphylococcus aureus

US-09-815-242-13158

Query Match 4.0%; Score 209.5; DB 9; Length 1002;

Best Local Similarity 19.2%; Pred. No. 8.5e-05;

Matches 206; Conservative 159; Mismatches 354; Indels 353; Gaps 56;

QY 9 LQRMVYKPRIOEY-----VAILNALEYHN-----MSSSVVEKYLKLDINLTDN 58

Db 120 IQK--YVKV---DGEWLYKLVSVIESIDEYQNLKVIWNSKAHIINDRIYL---NEQSAI 171

QY 59 YLNTYKSGRNKALKKFEYLWELKNSLTPVEKNLHFIWIGGOINDTAINYINQW 118

Db 172 YLNNKOTS-----FSIKNSL-----LVFIKTTVNNNTIFSULA 207

QY 119 KDVNSDYTVKVVDSNAPLINTLKKTIVESATNTLTSPRENLDPEFDYNK-----F 171

Db 208 EKINGEFEINDTKVYIHNVLVSKEIYSTTRPPL--SYSDNLN---YLNLKLSLHNDDF 263

QY 172 YKRMEIYYDKQHFIDYKSOIENPEFIIDNI-----IKTLYSNEYSKOLEALNK 223

Db 264 VKKIREI---QKLILAYEKTEIGFGEELYKDI IHMKALFKCKNYL---QIDTKIDMINN 317

QY 224 YIESLNKITTANGNDIRNLKFAEDLVRLNQBELVERWNLAAASDIURLIMLKEDGGV 283

Db 318 YLHQDI-----ATNISEAA-----YLLMLLSRNNIG----- 343

QY 284 YLDVDILPGIQDPLFKSINKPDSITNTSWEMIKLEAIMKYKEVIGYTSKNFMDLDE-- 341

Db 344 FIDLKVL-----HNRFEKYGFEOLVNIKOLISDITGFGTTIFQEEET 386

QY 342 -----VQSFESALSSKSKSEIFPLDDIKVSPLEVYKIAFANNSVINQALISLKDSY 394

Search completed: November 5, 2003, 19:41:22

Job time : 31.375 secs

Db 387 DGNNTVMLKQKFLHALRNN-----DEIVINEKQVE-SLINDNEINHYPMSADV 435

QY 395 CSDLVINOIKURYK--ILNDNLNPSINEGTDFNTMTKIP-SDKLASISNE----- 441

Db 436 YAEYLGRFYQYNELIVISPLTASFNAGATGFRPHHLIDTETLAKLBEKHHYQKMIC 495

QY 442 DNMMEMIKITNY-----LVKGFAPDVRSTINLSGFGVYTGA----- 477

Db 496 DONVEMISINIPKYPRNHNVLTHDSVEYSLNLG--SSNSYSKYELTLLDDIVGATFNKL 554

QY 478 --YQDLL-----MFKDNS-----TNIHLLEPELNFEPFKTK-ISQLTEQBITSL----- 519

Db 555 YLYSSQLNKRVLFESENMYNFKECNLYRL---LREISMESVKICEPMNDVSDISFSYSP 611

QY 520 -----WSPNQ-----ARAKSOFEEYKKGYPFEGALGEDDNLDFACNTVLD 558

Db 612 RIRYKQVILKPAYWKINEMVLPKPNVEWDQOFLKYQ-----EQFNININLV-- 660

QY 559 KDYVSKKILSSMKTRNKEYI-----HYIVLQGDKISYEASCNLFSKOPY 603

Db 661 --YGDNKLNLNLSLANHRYLLMKEYKHKRVLVESFLPQSKNDHV-YEIVTPIYKSSY 717

QY 604 SSILYQKNIEGSETAYYYVADAEIKEDKRIPIYQISNKRNIKLTFTGHGKSEFNTDTP 663

Db 718 -----RGPEI-----EIPKYK-----NTDIEYDKOMF 739

QY 664 A-NLDVDSLSSEIETILNAKADISPKYIEINLLGCNMFYSYIAEETYPGKLLKIKDR 722

Db 740 ALHIHIDKPSQDTFIIDNL-----YPFVKHLKDKGD 770

QY 723 VSE--LMPISIQDSITVSANQYEVRI--NEEGKREILDHSGKWINKESIIKDISKEYI 778

Db 771 IDQYFLMEYIKQGD-----LKLRLPRNDENYAEIYSLKNWLPVHRQ-TTEVSDYERFV 823

QY 779 SFNPKENKIIVKSKYLHELSTLLQEIRN---NANSSDIDLEKKVMLTEC-----EINVAS 830

Db 824 SYEPEFFRYGGKNT-INBIEAFFEYDTNLAVNIIENDPKFORPYIVAISIMYLFEMFESIS 882

QY 831 NIDR-QIVEGRIEBAKNLTSDSINYIKNEFKLIESISDSL-YDLAKHONGL-----DDSHF 883

Db 883 NEERMEIVNNYV--PTSPKSDIRPFKNELVTICNPANFEYMAKHYGIYRILKDKNQI 940

QY 884 IS--FEDISKTENGFRIRFINK---ETGNSIF-LETEKEIFSEYATHISKEI 929

Db 941 LSKLNEGLKTLTTKRSRIIGSLIHMRCNRIFGIDKQDQTF---VLSIVKEI 989



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:47 ; Search time 9.375 Seconds  
(without alignments)  
5116.506 Million cell updates/sec

Title: US-09-126-816b-6\_COPY\_1\_1020

Perfect score: 5192

Sequence: 1 MNLVNAQLQKVMYVKFRIQ.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4086	78.7	2366	1	TOXB_CLODI
2	2783	53.6	2710	1	TOXA_CLODI
3	293.5	5.7	1251	1	RBP2_PLAVB
4	273.5	5.3	2869	1	RBPI_PLAVB
5	257.5	5.0	1162	1	BXEN_CLOBO
6	256.5	4.9	1875	1	MLP1_YEAST
7	254	4.9	2136	1	YCF2_MARPO
8	249	4.8	1630	1	MSPI_PLAFK
9	249	4.8	1639	1	MSPI_PLAFW
10	247	4.7	1679	1	YIO9_YEAST
11	245	4.7	1701	1	MSPI_PLAFF
12	244	4.7	1682	1	MSPI_PLAF3
13	243	4.7	1701	1	MSPI_PLAFM
14	243	4.7	1726	1	MSPI_PLAFB
15	243	4.7	1726	1	MSPI_PLAFP
16	243	4.7	1928	1	MYSL_YEAST
17	237.5	4.6	2339	1	RPCI_PLAFA
18	237.5	4.6	3433	1	UTRO_HUMAN
19	237	4.6	1005	1	RA50_METJA
20	237	4.6	1163	1	SBCC_CLOAB
21	236.5	4.6	1790	1	USOI_YEAST
22	236.5	4.6	1957	1	SPOF_SCHPO
23	235	4.5	1179	1	EXSB_BUCAP
24	235	4.5	3210	1	CENP_HUMAN
25	234.5	4.5	1162	1	BXEN_CLOBU
26	232	4.5	1169	1	SMC_METJA
27	231.5	4.5	950	1	Y511_RICPR
28	231.5	4.5	1290	1	RA50_SCHPO
29	231	4.4	1225	1	Y309_MYCGE
30	228	4.4	1324	1	SMC4_SCHPO
31	227	4.4	1658	1	YMG7_YEAST
32	226	4.4	1803	1	YJL3_YEAST
33	225.5	4.3	1727	1	ALM1_SCHPO

RESULT 1

ID	TOXB_CLODI	STANDARD;	PRT;	2366 AA.
AC	P18177;			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Toxin B.			
GN	TOXB OR TCDB.			
OS	Clostridium difficile.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1496;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VPI 10463;			
RC	von Eichel-Streiber C.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
[3]				
RP	SEQUENCE OF 1271-2366 FROM N.A.			
RC	STRAIN=VPI 10463;			
RC	MEDLINE=922933124; PubMed=1603068;			
RA	Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J., Sauerborn M.;			
RT	"Comparative sequence analysis of the Clostridium difficile toxins A and B.;"			
RL	Mol. Gen. Genet. 233:260-268(1992).			
CC	-!- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN ENTEROTOXIN CALLED A AND CYTOTOXIN B.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X531138; CAA37298.1; -			
DR	EMBL; X92982; CAA63562.1; -			
DR	EMBL; X60984; CAA43299.1; -			
DR	PIR; A27636; A27636.			
DR	PIR; S10317; S10317.			
DR	InterPro; IPR002479; CW binding.			
DR	Pfam; PF01473; CW binding_1, 17.			
KW	Toxin.			
SQ	SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;			

Q04956 plasmodium  
Q58611 methanococc  
P38851 saccharomyc  
P52362 human herpe  
O51578 borrelia bu  
P33539 agaricus bl  
O14157 schizosacch  
P47460 mycoplasma  
P46081 clostridium  
P34036 dictyosteli  
P08799 dictyosteli  
Q92351 schizosacch

Query Match		78.7%; Score 4086; DB 1; Length 2366;
Best Local Similarity		77.5%; Pred. No. 9.1e-150;
Matches	790; Conservative	105; Mismatches 125; Indels 0; Gaps 0;
QY	1	MNLVNAQLOKMYVFRQEDYVAILNALBEYHNMSESSVVEKYKLKDINNLTNYL 60
DB	1	MSLVNRKQLEKMANVFRQEDYVAILDALBEYHNMSESSVVEKYKLKDINNLTNYL 60
QY	61	NTYKSGRKNALKKFKEYLTMTVELEKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB	61	DTYKSGRKNALKKFKEYLTMTVELEKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
QY	121	VNSDYTVKFDYDGNALINTLKTIVESATNTLTLESFRENLDNPEFDYNNFKYKRMELIY 180
DB	121	VNSDYVNVFYDGNALINTLKTIVESAINDTLESFRENLDNPEFDYNNFKYKRMELIY 180
QY	181	DKQKHFIDYKSIENPEFIIDNIIKTYLSNEYSKDLALNKYIEESLNKITTANNNDI 240
DB	181	DKQKNFINYKQAQREENPELIIDDIIVKTVLSNEYSKEIDELNTYIEESLNKITSQNGDV 240
QY	241	RULEKFADEDLVRLNQELVERNLAAASDILRISMLKEDGGYLDVLDLPGIQDPLFKS 300
DB	241	RNFEEFKNGESFNLYEQELVERNLAAASDILRISALKEIGGYLDVLDLPGIQDPLFS 300
QY	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORSPEALSCKSKSEIF 360
DB	301	IEKPSVTVDFWEMTKLEAIMKYKEYIPYTSSEHFDMLDEEVQSSPEVSLASKSKSEIF 360
QY	361	LPDDIKVPLEVKIAPANNVINQALISLKOSYCSDLVINQIKNRYKILNDNLNPSINE 420
DB	361	SSLGDMEASPLEVKIAFNKGIINOGLISVKOSYCSNLIVQIENRYKILNLSNPAISE 420
QY	421	GTDFTNTMTKIFSKLASISNEDNMFMWIKITNYLKYGFAPDYRSTINLSGPGVYTGAYD 480
DB	421	DNDFTNTTTFDISMAENADNGRFMELGKLYLRVGFDPDKVTITNLSGPAYAAAYD 480
QY	481	LMFMKDNSTNIHLLEPELNRFPKTKISQLEQETISLWSFNQARAKQFEYKKGYPE 540
DB	481	LMFMKSGSNHILIEADLNLFELSKTNIQSQTEQEWASLWSPDDARAKQFEYKKNYPE 540
QY	541	GALGEDNDLDFQAQTVLDKDYVSKKILSSMKTNKEYIHYIYVQLQGDKISYBASCNLFSE 600
DB	541	GSLGEDNDLDFQNIYVVDKEYLEKISSLARSSRGYIHYIYVQLQGDKISYEACNLFPAK 600
QY	601	DPVSSILYKQNIETSETAYYYVADAEIKEDYRIPYQISKNRKNKLTFFIGHGSEFNT 660
DB	601	TPYDSVLFOKNIBDSIAYYNPGDGEIQEIDKYKIPSIISDRPKLTLFFIGHGKDEFN 660
QY	661	DTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFYSIYAEETYPGKLLKIK 720
DB	661	DIPAGFDVDSLSEIEAIDLAKEDISPKYIEINLGCNMFYSINVEETYPGKLLKVK 720
QY	721	DRVSELMPSISQDSITVSANQYEVRIEENKGRILDHSGKWNKEESIIKDISKEYISF 780
DB	721	DKISELMPSISQDSIIVSANQYEVRIEENKGRILDHSGKWNKEESIIKDISKEYISF 780
QY	781	NPKENKIIVKSKYLHELSTLLOIRNANSDDIDLEKKVMTCEINVASNDRQIVGR 840
DB	781	NPKENKITVKNLPELSTLLOIRNANSDDIDLEKKVMTCEINVASNDRQIVGR 840
QY	841	IEBAKNTSDSINYIKNEFKLESISDLSYDLKHQGLDDSHFISPEDISKTENGPRIF 900
DB	841	IEBAKNTSDSINYIKNEFKLESISDALCDLQKQNELEDSHFISPEDISEDEGFSIRP 900
QY	901	INKETGNSIFIEFEKEIFSEYATHISKEISNKTOTIPDNVNGKLVKVNLDAAHVNTLN 960
DB	901	INKETGESIFVETEKIFSEYANHITEEISKIGTIFDTVNGKLVKVNLDTHVNTLN 960
QY	961	SAFFIOSLLEYNTKESLSNLSVAMKVQYVAQLFSTGLNTITDASKVVELSTALDETID 1020
DB	961	AAFFIOSLLEYNSKESLSNLSVAMKVQYVAQLFSTGLNTITDAAKVELSTALDETID 1020

RESULT 2	
TOXA_CLODI	STANDARD; PRT; 2710 AA.
ID	TOXA_CLODI
AC	P16154;
DT	01-APR-1990 (Rel. 14, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	Toxin A
GN	TOXA OR TCDA
OS	Clostridium difficile.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
OX	NCBI TaxID=1496;
RN	(1)_TaxID=1496;
RP	SEQUENCE FROM N.A.
RC	STRAIN=VPI 10463;
RX	MEDLINE=90221894; PubMed=2109310;
RA	Sauerborn M., von Eichel-Streiber C.;
RT	"Nucleotide sequence of Clostridium difficile toxin A.;"
RL	Nucleic Acids Res. 18:1629-1630(1990).
RN	(2)
RP	SEQUENCE FROM N.A.
RC	STRAIN=VPI 10463;
RX	MEDLINE=90129305; PubMed=2105276;
RA	Dove C.H., Wang S.Z., Price S.B., Pheiphs C.J., Lyster D.M.,
RA	Wilkins T.W., Johnson J.L.;
RT	"Molecular characterization of the Clostridium difficile toxin A
RT	gene.;"
RL	Infect. Immun. 58:480-488(1990).
RN	(3)
RP	SEQUENCE FROM N.A.
RC	STRAIN=VPI 10463;
RA	von Eichel-Streiber C.;
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC	-!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC	REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC	DIFERENT OLIGOPEPTIDES.
CC	-!- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC	ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
CC	DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
CC	CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
EMBL; X51797; CAA36094.1; -	
EMBL; M30307; AAA23283.1; -	
EMBL; X92982; CAA63564.1; -	
InterPro; IPR002479; CW binding.	
Pfam; PF01473; CW binding 1; 28.	
Pfam; PF04488; Gly_transf_sug; 1.	
Toxin; Enterotoxin.	
SEQUENCE 2710 AA; 308052 MW; 0A6B52CE84C14421 CRC64;	
Query Match	53.6%; Score 2783; DB 1; Length 2710;
Best Local Similarity	52.1%; Pred. No. 1.1e-99;
Matches	534; Conservative 218; Mismatches 265; Indels 8; Gaps 5;
QY	1 MNLVNAQLOKMYVFRQEDYVAILNALBEYHNMSESSVVEKYKLKDINNLTNYL 60
DB	1 MSLISKEELIKLAY-SIRPRENEYKTLTNDYKNTNNNNENKYLOLKKLINESIDVFM 59
QY	61 NTYKSGRKNALKKFKEYLTMTVELEKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB	60 NRYKTSRRNALSNLKKOILKEVILKNSNTSPVEKNLHFWIGVEVSDIALEYIKQWAD 119
QY	121 VNSDYTVKFDYDGNALINTLKTIVESATNTLTLESFRENLDNPEFDYNNFKYKRMELIY 180

Db 120 INAEYNIKLWYDEAFVNTLTKAIVESSTTEALQLEEEIQNPQDNMKPKYKGRMEFTY 179  
 QY 181 DKQKHPIDYKQGIENPEFIIDIIKTYLSNEYSKDLALNKYITEESLKNKTANNNGDI 240  
 Db 180 DRQKRPINYKQIQNKPTPTTDDIIKSHLSEYHNDETVLESYRTNSURKINSHHGDI 239  
 QY 241 RNLEKPADBLVRLNQELVERWNLAASDIILIRISMLKEDGGVYLDVILPGIQDPLPKS 300  
 Db 240 RANSLPTEQELLENIYSQELNLRGNLAASDIIVRLALKNFGVYLDVDMPLGHSIDLPKT 299  
 QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNPDMLEEVQVRSFESALSSKSKSEIFP 360  
 Db 300 ISRPSISIGDRWEMIKLEAIMKYKEYINNTYSENFDKLQDLQKNDPKFLIESKSEIFP 359  
 QY 361 LPLDDIKVPLEKVFAPANNVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420  
 Db 360 SKLENLNSDLAIKIFALGVSINQALISKQSYLTNLVIEQVKRYQFNLQHLNPAIES 419  
 QY 421 GTDFNTMTKIFSDKLASISNEEDNMFMPIKITNYLVKGFAPDVRSTINLSGPGYTTGAYOD 480  
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISLSGPGAYASAYD 479  
 QY 481 LLMFKDQNSTNIHLLPELNFPPPKTKISQLTEQETISLWSFNQAPAKSQFBEYKKGYPE 540  
 Db 480 FINLQENTIEKTLKASDLIEFKPPENNLSQLTEQEINLSWSPDQASAKYQFEKYVRDYG 539  
 QY 541 GALGEDNDLPAQNTVLDKDY-VSKKILSS--MKTENKEYIHYVLOQDKISYEASCNL 597  
 Db 540 GLSUEDONGVDFNKNTALDKNLLNKKIPSNVVEEAGSKNYVHYIQLQGGDISYEATCNL 599  
 QY 598 FSKDPYSSILYQKIEGSETAYYYVAD--ABIKEIDKYRIPIYQISNKENIKLTFIGHCK 655  
 Db 600 FSKNPKNSIIQRM--NESAKSYPLSDGSEILELANKYRIPERLKNEKVKVTFIGHCK 657  
 QY 656 SPFNTDTPANLVDLSLSSEIETILNAKADISPKYIEINLGCNMFYSIYAEETYPGKL 715  
 Db 658 DBFNTSEFARLSVDSLNSBISSFLDTIKLDISPKNVENVLLGCNMFYSDFNVEETYPGKL 717  
 QY 716 LKIKDRVSELMPISIQSDSITVSNQOYEVINEEGKREILDHSGKWINKEREIIDKISSK 775  
 Db 718 LUSIMDKITSTPDVKNKSITIGANQYEVIRINSEGRKELLASHGKWINKEREIIMDSLSK 777  
 QY 776 EYISFNPKENKIIVKSYHELSTLQEIIRNANSSDIDLEKKVMLTECEINVASNIDRQ 835  
 Db 778 EYIFPDSIDNKLKAKSNIPGLASISEDIKTLLLDASVSPDTKFIILNKLNIASSIGDY 837  
 QY 836 IVEGRIEAKNLTSDINYNKPEFKLIESISDLYDLKHONGLDSDHSFISPDISKTENG 895  
 Db 838 IYVEKLEPVKNIIHNSIDDLIDEFNLENNVSDELYELKLNLDKYLISFEDISKNST 897  
 QY 896 PRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHE 955  
 Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHTKEISTIKNSITIDVNGNLDNLDIDHTSQ 957  
 QY 956 VNTLASAPFIQSILIEYNTTKESLSNLSVAMQVYAQLFSTGLNTITDASKVVELVSTAL 1015  
 Db 958 VNTLNAAPFIQSLIDISSKNQVLDNSTSVKQVLYAQLFSTGLNTIYDLSIQLVNLISNAV 1017  
 QY 1016 DETID 1020  
 Db 1018 NDTIN 1022

## RESULT 3

DBP2\_PLAVB STANDARD; PRT; 1251 AA.  
 AC Q00759;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Reticulocyte binding protein 2 (fragment).  
 GN RBP2.  
 OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 RN NCBI\_TaxID=31273;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92315338; PubMed=1617731;  
 RX Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 merozoites";  
 RL Cell 69:1213-1226(1992).  
 CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 HUMAN RETICULOCYTE CELLS.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M88098; AAA29744.1; -;  
 KW Malaria; Receptor; Membrane.  
 FT NON\_TER 1  
 FT NON\_TER 1251  
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 5.7%; Score 293.5; DB 1; Length 1251;

Best Local Similarity 19.7%; Pred. No. 0.00018;

Matches 207; Conservative 194; Mismatches 405; Indels 247; Gaps 45;

QY 119 KDVSNDYTKFYVDYDNALFINTLKTIVESATNNTLSFRENLDPEFDYKPY----- 172  
 Db 8 KDTSEDEKKSIEKAYEKMGNTLK-----ELEKDDKEKIEKEVEEAQIQKFIPIHDVN 63  
 QY 173 -----RKRMELIYDKGFIDYKQIENPEFIIDIIKTYLSNEYSKDLBALNKYIEE 227  
 Db 64 LMNDEVEKSKIVWEK-----IELYKEIDEIKQ-----KT--NEY-KQGDTSNFYITE 108  
 QY 228 SLANKITANNNDIRNLEKFADEDIVRLYNQELVERWNLAASDILRISMLKEDGGVYLDV 287  
 Db 109 QYNSATQSKA-----KIQFI-----NIATTK-----KQTSQSDI 140  
 QY 288 DILPGIQDPLFKSINKPDSITNTSWEM-----IKLEAIMKYEIPGYTSKN----- 334  
 Db 141 NELESIEKEVHKNLQLVKQESNMEEMRKQILSMKOLLINSETIAKEISNTQALCF 200  
 QY 335 -----FDMLDEEVQVSFESALSSKSKSEIFLPLDDIKVSPLEVKIAPANNVINQA- 386  
 Db 201 RENAKTKLNKTDELLQVAAIEEAKAHKNIDIALEDAQIDTEVSKIEQINREIMNKKD 260  
 QY 387 -----LISLKD--SYCSDLVINQIKRYKI-----LNDNLNPSI 418  
 Db 261 EIKSYLSIEKYEKDKCTTEISNSKRGKQIEFKPEKPNESNKNVINEINENIRNSE 320  
 QY 419 NEGTFD-----NTTMKIFSDKLASISN--EDNMFMFKITNYLVKGFAPDVRSTINLS 469  
 Db 321 QYVKDIEDAKQASQKVELFKHETIISNIPKESILGVETSKQKINKAEIMKEIERH 380  
 QY 470 GPGVYT--GAYQDLLMFKDNSTNIHLLPELNFPPPKTKISQLTEQEITSLWSFNQARA 527  
 Db 381 NSEIQTVKGFQENLKNLNEPHYDNEDELNDKSTNAKV--LIETNLESV-----KHN 433  
 QY 528 KSQFEYKKGYPEGALGE---DDNLDLPAQNTVLDKDYVSKKILSSMKNKTKYIHYIQL 584  
 Db 434 LSEITNIKQG-----GEKIYSKARDIMQKIKATSENTAEKTEKVKDQSQSNVNYL--- 484  
 QY 585 QGDKISYEASCNLFKSDPYSSI-LYQKNIEGS--ETAYVYVADAE-IKEIDKYRIPYQI 640  
 Db 485 --NQITERNLIVTEKRLNGIDSTITNIEGALKESKGYEIGFLEKLEIGKRLKVD 542  
 QY 641 SNKRNIKLTFFIGHKSEFNT-----DTFANLD-----VDSLSEIETITLNLAKADI--- 696  
 Db 1018 NDTIN 1022

543 ITKKSINST-VGNFSSLNFDLNDQYFNKNINDYENKMGIEYNEPEGSLNKISENLRNA 601  
 687 QY -----SPKYL-----EINLG-----CNMFYSIYAEYTPKLLKIKDRVSELM 727  
 602 DB SENTSYSAKTLURLEAQKEKVNLLKNEEANKYLRDVKKVESF--RFIFNMKESLDKIN 659  
 728 QY PSISQSDITVSANQYVRINEEGRREILDHSGKWINKESIIKDISKEYISFNPKENKI 787  
 660 DB EMKESQITVNEGHNGVQVLEIKELVDE-----NNLSDLAQATGKN-----BEIQ 708  
 788 QY IVKSKYLHELSTLLOIRNANSDDILEKKVMLTE-----CEINVASNIDRQIVEGRIBE 843  
 709 DB ITHSTLKNRAKTLGHVDFSAKVGIKITPELALTELLGDAKLKTAQELKFESKNNVLE 768  
 844 QY AKULTSDSINYIKVEKFLIESISD-----SLYDLKHONGLDDSHFISFEDISK-TENG----- 895  
 769 DB TENMSKNT-----NELDVHKNIODAYKVALEILAHSDIEDTKQ-----KDSKLIEMGNQI 819  
 896 QY -PRIRFNKETGNSIFETEKEIFSEVATHISK---EISNI---KDTIFONVNGKLVKYN 949  
 820 DB YLKVVLINQYKNKISSKKEEAVSVKIGNVSKHSELSKITCSDSYNI-----TA 872  
 950 QY LDAAEHVNTLNSAFFIQ-----SLIEYNTTKESLSNLSVAMK----- 986  
 873 DB LEQTELQNLNRSFTQETKNTNSDSKLEIKITDFESLKNALKTLEGEVNAKASSDNHEH 932  
 987 QY VOYQAQFSTGLNTITDASKVVELSTALDETI 1019  
 933 DB VQSKSEPVNPALSEIEKEETDIDSLNLTALDELL 965

RESULT 4

RBPI\_PLAVB STANDARD; PRT; 2869 AA.  
 ID RBPI\_PLAVB  
 AC Q00788;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Reticulocyte binding protein 1 precursor.  
 GN RBPI.  
 OS Plasmodium vivax (strain Belem).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=31273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92315338; PubMed=1617731;  
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RT "A reticulocyte-binding protein complex of Plasmodium vivax  
 merozoites";  
 RL Cell 69:1213-1226(1992).  
 CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 CC HUMAN RETICULOCYTE CELLS.  
 CC -I- SUBUNIT: Homodimer (Potential).  
 CC -I- SUBCELLULAR LOCATION: Membrane-bound.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M89097; AAA29743.1; -  
 DR Malaria; Receptor; Signal; Transmembrane.  
 KW SIGNAL 1 17  
 FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.  
 FT DOMAIN 18 2807 EXTRACELLULAR.  
 FT TRANSMEM 2808 2826 POTENTIAL.  
 FT DOMAIN 2827 2869 CYTOPLASMIC.  
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 5.3%; Score 273.5; DB 1; Length 2869;  
 Best Local Similarity 20.2%; Pred. No. 0.0028;  
 Matches 245; Conservative 210; Mismatches 408; Indels 351; Gaps 63;  
 6 KAOLQKQVVKPRIODEY-----VAINALAEYHNMSSESSVVEKYKLK-DINN- 55  
 946 KALKKIVSDSLRDKIDQYEFTEFKTSVAVENTVSTIQSLKA--IDSLKRLNGSINNCK 1003  
 56 ---TD-NYLNTYKSGRNKALK-----KFKEYITMEVLELKNLSLTPVEKNLHFIWIG 104  
 1004 KYNTDIDLLRSKIKTLREEVKEMPKRGDKCGENTALLLLSLRDKMGKINEKLN---D 1059  
 105 GOIN--DT-----AINVINQKDV-----NSDYTV- 127  
 1060 GRNLSLDTKKEDLLKPYSESKEIHLKDKQGPQPLARIDEDIKRDVDELNVNQVI 1119  
 128 -----KFVYDSNAFL-----INTLKKTIVESATNNTLESFRE-----NLNDPFDYNK 170  
 1120 SENKVTLFQNSVTYIEAMHSHINTVAHCIT--SNKNEILKSVKEVEDKLNLAQEODEYK 1178  
 171 F---YRKMEII---YDKQKHFIDYKSOIBENPEFIDNIKIYLSN-----EYSKDLE 219  
 1179 VKNPENKQLEAIRGSMKSLKEVINKHVSEMTQ-----LESTANTLKSNAGKENEHDL 1233  
 220 ALNKY-----LEESLANKI-----TANNGNDIRLEKFADEDLVRLNQELVERWNL 266  
 1234 ELNKTGQWRDIYEKLLKIAELKEGTVNELNKANEKANKEPEPERNIIGHVLE----- 1288  
 267 AASDILRISMLKEDGGVVDILPGIQDLPKFSINKPDSITNTSWEMIKLEAIMKYKEY 326  
 1289 -----RITVEKDAGKV--VEEMSLKTKIEKLQ-----TSDDSQNELVTTISIKHLEN 1337  
 327 IFGYTSKNFMDLDEEVQSFESALSCKSDSEIFLPLDDIKVSPLEVKI-----AFANNV 382  
 1338 AKGY-----EDVIKRNEEDSIQLR-EKAKSLETLDEMCKLQVQVNMNLSAQGNAG 1388  
 383 INQALISLKDSCDLVINQIKNRYKILNDNLNPSINGTDTNTMTKPSDKLA----- 436  
 1389 ISKEL-----NELKGVIELLSTNYSSILEYVKNKNSVRFSQNGEFTK 1435  
 437 STSNEDNMFMKIKITNLYKVGPAPOVRSTINLSGPGVYTGAYQDMLLMFKDNSTNHLRP 496  
 1436 ASGEENKASARLAEKLEQIVKDL-----DYSN-----IDD 1468  
 497 ELRNFEPFKTKISQITQEBITSLNPFQNRASQOEYKGYFEGALGDDNLDFAQN-- 554  
 1469 KVKKIEGIRKILKMKESALT-FWBESE-----KFKQMCSSHMENAKEGKKIYELKNG 1522  
 555 -----TVLDK-----DYVSKILSSMKTRNKEYIHYIVLOQDKISYEASC----- 595  
 1523 DGGKANITDSQMEEVGNVYVK-----AEHAFHTVEAQVDKT--KAFESIVAYVT 1570  
 596 ---NLFSKDPYSSILYQKNIEGSETAYYYYVADAEIKEDKIRPYQIS-NKRNI----- 646  
 1571 KMDNLFNESLMKEVKVCKEKKNDAAEKY---SALKPYDG-RIKARVSENERKISELKE 1625  
 647 KUTFTIGHGKSEPN-TDTFANLDVDSLSEIETILNLAADISPKEYIENLLGCMPSYSI 705  
 1626 KAKVEKSSQLNDVSTKSLQIDNCRQQLDSVLS-----NIGRVQNALQYFD 1674  
 706 YAEETYPGKLLKIKDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINK 765  
 1675 SAKSKMSKSVL-----PISELGAESLDKVKAAKSEYKQL-----ETVQEMSRINVE 1722  
 766 ESIIDKISSKEYISFNPKENKIIIVKSKYLHELSTLLOIRNANSDDILEKKVMLTECE 825  
 1723 EGSLLTDIDKK-----ITDENDLLKMKQYEE--GLLOKIKENADKESNFE-----LVGSE 1772  
 826 INVASNIDRQI-VEGRIBEAKNLTSDSINY-----IKNEFKLIESISDLSYDLKHQNGL 878  
 1773 INALLDPSTSIPIKLKLE-YDMTGLDKNYGVKNMEIHGEF-----TKSYNLIETHLSNAT 1827

QY 879 DSHFISFE-----DISKTENG-----PRIRFINKETGNSIFIETKEIFSEYA 922  
 DB 1828 DYS--VTPKQAQSLRELAEKBEHLRRREAEALFLINDIKKVESLKLKEMMKVSAEYE 1885  
 QY 923 -----THLSKEISNKTIFPNVNG-----KLVKV-----NLDAAHEVN 957  
 DB 1886 GMRDHTSVSQLVQDMK-TIVDELTKTLNDISECSSVLNNVVSVIVKVKESKHADYRRDAN 1944  
 QY 958 -----TNSAFF-----IQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLTITDAS 1005  
 DB 1945 SMYESWTLNVLNPLSDEAKISSGMEFNA--EMKSNFKTLELEIP-----SVISNSN 1994  
 QY 1006 KVVELVSTALDETI 1019  
 DB 1995 ELLKKIEQDSNDVI 2008  
 RESULT 5  
 BLEN\_CLOBO  
 ID BLEN\_CLOBO STANDARD; PRT; 1162 AA.  
 AC P46082;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Botulinum neurotoxin type E, nontoxic component.  
 GN ENT-120.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Washington;  
 RX MEDLINE=93195515; PubMed=8450310;  
 RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tezuka K., Oguma K.;  
 RT "The complete nucleotide sequence of the gene encoding the nontoxic  
 component of Clostridium botulinum type E progenitor toxin."  
 RL J. Gen. Microbiol. 139:79-86(1993).  
 CC -!- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN  
 TOXICITY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D12697; BAA02194.1; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; Bontoxilysin.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 KW Neurotoxin.  
 SQ SEQUENCE 1162 AA; 136856 MW; 96468DDEDAE0F39D CRC64;  
 Query Match 5.0%; Score 257.5; DB 1; Length 1162;  
 Best Local Similarity 20.3%; Pred. No. 0.0039;  
 Matches 233; Conservative 160; Mismatches 424; Indels 331; Gaps 52;  
 QY 34 YHNSSESSVVEKYLKLD-----INNLTDLNLTYYKKSGRNKALK-- 73  
 DB 109 YENNTEDYQNTYLSKNNEHYHTANLVIFPGSGNIKNNVIYKKEAYESGGMGTWLEIW 168  
 QY 74 -----KFKELYTMVELEL-----KNNSLTPVPEKNL 98  
 DB 169 FQPLTHKYDEFYVDALEIKLSLYLYGIKPNNDNLNIPYRLRNEFNSLEYSELNW 228  
 QY 99 HFIWIGQNDTAINVINO-----KOVNSDYTKFVVDNSAFNLTKTKTI 145  
 DB 229 IDFLISGGIDYKLLNTNPYWFIDKYFDTSKNPEKYNDEYIKI--KNNNYIANSIKLYL 286

RESULT 6  
 MLPL YEAST  
 ID MLPL YEAST  
 AC Q02455;

STANDARD; PRT; 1875 AA.





DT	01-MAR-1989	(Rel. 10, Created)
DT	01-MAR-1989	(Rel. 10, Last sequence update)
DT	28-FEB-2003	(Rel. 41, Last annotation update)
DE	Hypothetical 259.9 kDa protein ycf2 (ORF 2136).	
GN	YCF2.	
OS	Marchantia polymorpha (Liverwort).	
OC	Chloroplast.	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;	
OC	Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;	
OC	Marchantiaceae; Marchantia.	
OX	NCBI_TaxID=3197;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ohyanono K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,	
RA	Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,	
RA	Ozeki H.;	
RT	"Chloroplast gene organization deduced from complete sequence of	
RT	liverwort Marchantia polymorpha chloroplast DNA.";	
RL	Nature 322:572-574 (1986).	
CC	-1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcement">http://www.isb-sib.ch/announcement/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; X04465; CAA28078.1; --	
DR	PIR; S01591; A05037.	
DR	InterPro; IPR003593; AAA_ATPase.	
DR	InterPro; IPR003959; AAA_ATPase_cent.	
DR	Pfam; PF00004; AAA; 1.	
DR	SMART; SM00382; AAA; 1.	
KW	Chloroplast; Hypothetical protein.	
SQ	SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;	
Query Match 4.9%; Score 254; DB 1; Length 2136;		
Best Local Similarity 21.1%; Pred. No. 0.011;		
Matches 228; Conservative 152; Mismatches 329; Indels 374; Gaps 55;		
QY	27	ILNALBEYHNMSSSVVEKYLKLDINNL-TDNYL-----NTYKSGRNKALKKPKLEYLT 80
DB	309	IFKNLQNF-NESDKLIESFLLKIKGLYFKNVIEPVWQSYKKQCLD--FNKFNELNN 365
QY	81	MEV-LELKNLSLTPVEKLNLFPIWGQINDAL-----NYINQWKDVS--DYTVKFVYDS 133
DB	366	SEYIKIEELFSYIYKFSYILYEGKSKTIIKQSFNNNIYYKKLNSFNFTIFYFDS 425
QY	134	NAFLINTLTKKTIVESATNNT--LESRENLDPEFDY---NKFYRKRMEIIVDKQKHF 187
DB	426	NNLFWLWKQNY--INNKPPLKSP-----LYSSISNQF-----ILFPKQKNSK 468
QY	188	DYYSQIEENPEFIIDNIITKYLs---NEYSKDL-----EALNKYIEESLNKIT 233
DB	469	SFNKNLVKKNSKDVITVPSKENKIEINNPSKSIYVAFPILSINEIDNKFV--INKIS 525
QY	234	ANNNDIRN---LEKFADEDLVRLYNQELVERVN-----LAAASDIIRIS-ML 277
DB	526	LKNINKKKQKRFYLNKIKSSDNFRFNLWKIKNYSQQFVSNNNSFLNPAFELQQNYL 585
QY	278	KEDGGVYLDVDILPGIQDLFKSINKPDS-----ITNTSWEMIKLEAMK--- 322
DB	586	KKKNIILF-----FKKLNEVFSNFFFYQYKCKKLNIFLKFALEKILKGRNK 632
QY	323	-----YKEYIPGYSKTNFMDLDEEVQSFSAKSKDSKEIFLPLDDIKV----- 368
DB	633	KFTISIKLFPKKYKKNLNGEYKIESQILQNEKLNKKRKKQFPNP--NIKLSFPNS 690
QY	369	SPLEVKIA-----PANNSVINQALLSLKDSYCDLVINQIKRYKILNDNLNPSINEGTFDN 425
DB	691	SKGNIYLNQYFNFKNLINNKLTWK-----KISNKLVISSEYN----- 730

Qy	426	TTWKIPSDKLAISNEDNMFMKI	NYLKVGPADPVRSTRINISGPGVYTGAVQD	LLMFK	485		
Db	731	---KI	---IWNNKMKF	---FSFKNSVL	---DTFFN	756	
Qy	486	DNSTNIHLPELNRNFEFPKTI	SOLTEQIITSLWS	---FNOARAKSOFEEYKKGYP	540		
Db	757	KKSFNIITV	---IFDKKKIQLNQFQIKLNC	FLSFLPFSKNIK	---KTKIFONSYP	807	
Qy	541	GALGEDNDLDAQNTVLDKDY	VSKKILSSMKTNRKEYIH	IVOLQSDKISYEASCNLF	FSK	600	
Db	808	---	---INENLTTFSPNDKEFN	IPFLFLFISEINNDFLMRFFK	846		
Qy	601	DPYSSILYQKNISGSETAY	YYYVADAEI	---KEIDKYRIPYQISKNRNIK	LTFTIGHGKSEF	658	
Db	847	---	---YLYYRIYKDKIELFPN	PIENRQLQNFPEKTKI	---LTFI	884	
Qy	659	NTDTFANLVDLSLSBIET	ILNLAKADISPKYTEINL	LCGMFYSYIABETYPGKL	LLK	718	
Db	885	LQD	---PELNNRPFIFHLEKTK	NN	---YLRLLKIP	919	
Qy	719	IKDRVSELMPISIQDSIT	VSANQYVRINSEKREILD	HSGKWINKKEESI	IKDISKEYI	778	
Db	920	LKDKRNFL	---	INE	---IK	933	
Qy	779	SFNPKNKIIVSK	---	---YLHELSTLLOETRNNANS	DDIDLEK	817	
Db	934	SFIEKKNNLFIKSQLSN	VLNVKNSYFPDFN	IFNPHFLKQKEKNI	IEILNNQNYPEKSL	993	
Qy	818	KVMLTECEINVA	---SNIDRQIVEGRIEAK	NLSDSINYIKNEFKLIES	ISDSLY	871	
Db	994	KTYLKNLNNYSKESYK	IF	QLNLN	---KNYKTFQWISLFI	YSKNLN	1044
Qy	872	LKHQGLDSDHFTSPED	ISKTENGFR	---IRPINK	---ETGNSIPFTEKEIFSEY	ATHIS	926
Db	1045	YKIQNKIEKNYCNKNSY	KKKKIKTVNPEFKNN	LFQTNNSWFFTL	---	WWEYNTYIL	1101
Qy	927	KEISNLIKOTIFDNV	---	---CKLKKVKNLDAAE	VNTLNSAFFQSLIEYNTTKES	LS	979
Db	1102	LQI	---IQETFFQITDVL	EVFKKKIIEK	---	NLKFPLKS	1141
Qy	980	NLS	982	---	---	---	---
Db	1142	TLS	1144	---	---	---	---
RESULT 8							
MSPI_PLAFK							
ID	MSPI_PLAFK	STANDARD;	PRT;	1630	AA.		
AC	P04932;						
DT	13-AUG-1987	(Rel. 05, Created)					
DT	01-FEB-1996	(Rel. 33, Last sequence update)					
DT	01-OCT-1996	(Rel. 34, Last annotation update)					
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)						
DE	(P04932) (P190).						
GN	MSP-1.						
OS	Plasmodium falciparum (isolate K1 / Thailand).						
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.						
OX	NCBI_Taxid=5839;						
RP	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=66136024; PubMed=3004972;						
RA	Mackay M., Goman M., Bone N., Hyde J.B., Scaife J., Certa U.,						
RA	Stunnenberg H., Bujard H.;						
RT	"Polymorphism of the precursor for the major surface antigens of						
RT	Plasmodium falciparum merozoites: studies at the genetic level.";						
RL	EMBO J. 4:3823-3829 (1985).						
RL	[2]						
RP	REVISIONS, SEQUENCE FROM N.A.						
RA	Pan W., Tolle R., Bujard H.;						
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.						
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor						
CC	(Potential).						

CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X03371; CAA27070.1; -  
CC InterPro: IPR006209; EGF\_like.  
CC Pfam: PF00008; EGF\_1.  
CC Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
CC Transmembrane; GPI-anchor.  
CC FT SIGNAL 1 19 POTENTIAL.  
CC FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.  
CC FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.  
CC FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.  
CC FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC SQ SEQUENCE 1630 AA; 187289 MW; ADBDE3C5E0A46322 CRC64;  
  
Query Match 4.8%; Score 249; DB 1; Length 1630;  
Best Local Similarity 20.9%; Pred. No. 0.012;  
Matches 257; Conservative 215; Mismatches 418; Indels 338; Gaps 69;  
  
QY 16 KPIQDEY-VAILN-ALEEYHNMSSVVEKYL-----KLKDINNITDNLNTYK----- 64  
DB 266 KKLYQAQYDLSYVQLQEAHLN--ISVLEKRIIDLTKKNEIKELLDK-INEIKNPPPA 322  
QY 65 KSG-----RNKAL-----KKFEY-----LTWEVLEL-----KN----- 88  
DB 323 NSGNTPTNLDKKNKIEEHEKEIKETAKTIKFNIDSLFTDPLEYLYLREKKNIDISAK 382  
QY 89 -----NSLT-PVEKHLFTWIGGQINDTAINYNOWKDVNS--DYT---VKF 129  
DB 383 VETKESTEPNEYPNGVTYPLSYN-----DINN-ALNELSFGDLINPFDYTKPESKN 433  
QY 130 VYDSN-AFLINTLKTKI-----VPSATNNTLESFRENLDNPEFDYKFKRMELIYD 181  
DB 434 IYTDNERKKEFEINEIKIEKKIEKIES-DKKSVEDRSKSLN-----DITKEYEKLNEIYD 488  
QY 182 -KQKHFDY--YKSOEENPEFIDNII--KTYLSNEYSK-DLEALN---KITEE-SINK 231  
DB 489 SKFNNIDLTNFKMGKRYSYVEKLTHNTTASYENSKHNLEKLTALKYKYMEDYSLRN 548  
QY 232 ITANN-----GNDIRNLEKFADEDLRLYNQELVERWN-----LAAASDILRI 274  
DB 549 IYVEKELKYKNLSKIENEIEITLVENIKDEQLPEKTKIDENKPEKILEVSDIVKV 608  
QY 275 SM-----LKEDGGVYLDVILPGIQ-PDLFKSINKPDSITNTSWEMIKLBAMK 322  
DB 609 QVQKVLMMNKIDELKKTQLIKNVELKHNHVPNSYKQENKQRPY-----YLVLKKEIDK 664  
QY 323 YKSYIPCYTSKFMFLDEE---VQSFESALSSEKSKSEI-----FLPLD 365  
DB 665 LKVFMPKVES-----LINEKKNTKTEGQSDNSPSTEGEITGQATTKPQQAGSALGDS 720  
QY 366 IKVS-----PLEVKI-----AFANNSVINQALISLKSYSCLDVLINQIKN 405

DB 721 VOQAQEQKQAOQPPVPVPEAKAQVPTPPAPVNNKTENVSKLDYLEKLYEPLNTSYICH 780  
QY 406 RYKILNDNLNPSINEGTFTNTMTKIFSDKLASISNEDNMFMKITYNLVKGFADVRST 465  
DB 781 KYILVS---HSTMNE--KILQYKITKEESKSLSCDPLDLLFNITQNNIPVMYS--MFD 833  
QY 466 INLSGPGVYTGAYQ-----DILMFKDNSTNIHLPELNFEPFKTKIS-----QLTEQ 514  
DB 834 LNNLSUQLFMETIEYKEMVCNLYKLDNDKIKNLEBAKKVSTSVKLTSSSSMQPLSLTPQ 893  
QY 515 EITLSWSFNQARAKSQFEYKGYEGEGALGEDNDLDFAQNTVLDK--DYVSKKILSSMKT 572  
DB 894 DKPEV-SANDTSHSTNLNLSLKLPENLISLGNKNYIQLIGQSSSENFYEKILKDSDT 952  
QY 573 RNKEIYHIVOLQDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYYVADAEIKED 632  
DB 953 FYNESFTNFVKSADDIN---SLNDESK-----RKKLE-----EDIN 986  
QY 633 KYRIPYOIS---NKRNIKLTFIGHKSEFNTDTFANLDVDSLSSEIETILNLAKADISP 688  
DB 987 KUKTKLQSLFDLYNKYKLERLFDKKTGVGYKMKQIKKLTLLKEQLESKLSLN---NP 1043  
QY 689 KYIEINLLGCNMFYSIY-----ABETYPG-KLLAK-----IKDRVSELMPP--S 729  
DB 1044 KHV-----LQNFVFFNKKKEAEIAETENTLENTKLLKHGYKLYKYNNESSPLKT 1095  
QY 730 ISQDSITVAN-----QYEVRIINEGK-----REILDHSGKWINKESIIKD-ISK 775  
DB 1096 LSEESIQTEDNYASLENFKVLKSLGKGLKLNLEKKLSYLSGLHHLIAELKEVIRNK 1155  
QY 776 EYISNPKENKIIVKS-----KYLHE---LSTLQEIERNANSSDIDLEKKVMLTEC-- 824  
DB 1156 NYTGSPSENNTDVNNALESYKFLPEGTDVATVYSE--SGSDTLLEQSQPKAPASTHVA 1213  
QY 825 ---EINVASNIDRQI-----VEGRIBEAKNLTSDSINYIKNEFKLIESIDSPLYDLKH 874  
DB 1214 ESNTITTSQNVDDVDVLIPIFGESE---DYDDLQGVVTGEAVTPSVIDNI----- 1264  
QY 875 QNGLDDSHFISFEDISKTEGFRIRINKEGNSIFITEKEIFSEYATHISKEISNIKD 934  
DB 1265 -----LSKIENEYEVLYKPLAG--VYRSLKKQLENNVMTF-----NVNVN 1304  
QY 935 TI-----PDNV-NGKLVKKNLDAAEV-----NTLASAFFIQSIJEYNTTKESL- 978  
DB 1305 IUNSFNKNRENKPNVLESDDLPIYKOLTSNYYVVKDPYKFLNKEKDKFLSSNYNKDSID 1364  
QY 979 SNLSVAMKVQVYAQLFSTGLNITITDASK 1006  
DB 1365 TDINPANDVLGYKYLSEKYSKDLDSIK 1392  
  
RESULT 9  
MSPI\_PLAFW  
ID MSPI\_PLAFW STANDARD; PRT; 1639 AA.  
AC P04933;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMWSA) (P195).  
GN MSP-1.  
OS Plasmodium falciparum (isolate Wellcome).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5848;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86014355; PubMed=2995820;  
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,  
RA Nicholas S.C., Hillman Y., Davey L.S., Iizard M.L.V., Schwarz R.F.,  
RA Freeman R.R.;  
RT "Primary structure of the precursor to the three major surface  
RT antigens of Plasmodium falciparum merozoites.";

RL Nature 317:270-273 (1985).  
RN (2)  
RP REVISIONS.  
RA Holder A.A.;  
RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
CC (potential).  
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
CC kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X02919; CAA26676.1; -  
DR PIR; A24594; A24594.  
DR PIR; S05603; S05603.  
DR PDB; 1CEJ; 28-MAY-99.  
DR InterPro; IPR006209; EGF\_like.  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor; 3D-structure.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;  
  
Query Match 4.8%; Score 249; DB 1; Length 1639;  
Best Local Similarity 20.9%; Pred. No. 0.013;  
Matches 257; Conservative 215; Mismatches 418; Indels 338; Gaps 69;  
  
QY 16 KFRIOEDY-VAILN-ALBEYHNMSESSVVEKYL-----KLKDINNLTNLYNTYK----- 64  
DB 275 KKKLYQAQYDLSYNNKQLEAHNL--ISVLEKRIDTLKKNENIKELLDK-INEIKNPPPA 331  
QY 65 KSG-----RNKAL-----KPFKEY-----LTMEVLEL-----KN----- 88  
DB 332 NSGNTPTLLDKNKKIEHEKEIKETAKTIKFNIDSLFTDPLEYLEYLRKKNKIDISAK 391  
QY 89 -----NSLT-PVEKNLHPIWIGGOINDTAINYNQWQVNS--DYT-----VKF 129  
DB 392 VETKSTENPENGVTYPLSYN-----DINN-ALNELNSFGDLINPFDTYKPSKN 442  
QY 130 VYDSN--AFLINTLKXTI-----VESATNTLTSPRENLDNPFDPYKFKRMEIYD 181  
DB 443 IYTDNERKKFINEIKIEKIEKKIES-DKSYEDRSKSLN-----DITKEYEKLNEIYD 497  
QY 182 -KQKHPIDY--YKSOITEENPEFIIDNI--KTVLSNEYSK-DLEALN-----KYIEE-SLAK 231  
DB 498 SKFNNNIDLTNPEKMGKRYKYVEKLTHHTNTPASYENSKHLEKLTALKATKYMEDYSLRN 557  
QY 232 ITANN-----GNDIRNLEKFADEDLVRLYNQELVERWN-----LAAASDILRI 274  
DB 558 IYVEKELKYKNKLNISKIENIEITLVENIKKBEQLEFKEKTKIDENKDPDKILEVSDIVKY 617  
QY 275 SM-----LKEDGGVLDVDILPGIQ-PDLFKSINKPDSITNTSWEMIKLEAIMK 322

DB 618 QVQKVLNKKIDELKKTQLILKNVELKNIHVPSYKQENKQEPY-----YLVLKKEIDK 673  
QY 323 YKEIYIPGYTSKNFDMLEDE--VQSPESALSSKSDKSEI-----FLPLDD 365  
DB 674 LKVFMPKVES-----LINBEKKNIKTEGQSDNSEPSTEGEITQCATTKPGQQAAGSLEGDS 729  
QY 366 IKVS-----PLEVKI-----AFANNVINQALISLKDYSYCSDLVINQIKN 405  
DB 730 VQAQAQEQKQAQPPVPVPPEAKAQVPTPPAPVNNKTNVNSKLDYLEKLYEFLNTSYICH 789  
QY 406 RYKILNDNLNPSINRGTDFTNTMKIPSDKLASISNEDNMFMKITYLVKGFADPVRST 465  
DB 790 KYILVS-----HSTMNE--KILQYKITKEESKLSKCDPLDLNFIQNNIPVMYS--MFDS 842  
QY 466 INLSGPGVVTGAYO-----DILLMPKDNSTNHLLEPELRNPFEPKTKIS-----QLTQ 514  
DB 843 LNNLSQJFMIEYEMVCMYKLDNDKIKNLLBEAKKVSVTSLSSSSQPLSLTPQ 902  
QY 515 EITSLWSFNQARAKSQPBKYGYPGEGALGEDNDNLDFAQNTVLDK--DYVSKILSSMKT 572  
DB 903 DKPEV-SANDDTSHSTNLNLSLKLFENILSLGKNKIYQELIGKSSSENFYKILKDSPT 961  
QY 573 RNKEVIHVIVOLGDKISYEASCNLFSDPYSIYQKNIEGSETAYYYYVADAIEKID 632  
DB 962 FYNESFTNFVKSADDIN--SLNDESK-----RKKLE-----EDIN 995  
QY 633 KYRIPYQIS-----NKENIKLTFIGHGKSEFNTDTPANLDDVDSLSSEIETILNLAKADTSP 688  
DB 996 KLKTKLQSLFDLYNKYKLERLFOKKVTGVKYNQIKKLTLLKQLESKLSLN---NP 1052  
QY 689 KYIEINLLGCNMFYSIV-----AETYPG-KLLK-----IKDRVSELMW--S 729  
DB 1053 KHV-----LQNESVFNNKKEAEIAETENTLENTKILLKHYKLVKYNGESSPLKT 1104  
QY 730 ISQDSITVSAN-----QYEVRIINEGK-----REILDHSGKWKINKEESIKD-ISK 775  
DB 1105 LSEESIQTEDNYASLENFVKLSKLEKLNLEKLSYLSGLHLHLIAELKEVKNK 1164  
QY 776 EYISFNPKENKIYKS-----KYLHE--LSTLQEIERNANSSDIDLEKKVMLETC-- 824  
DB 1165 NYTGNPSENNNTDNNALYESYKFLUPEGTDAVTAVYSE--SGSDTLEQSQPKPASTHVA 1222  
QY 825 ---EINVASNIDROI-----VEGRBEAKNLTSDSINYIKNBEFKLIESIDSYDLKH 874  
DB 1223 ESNTITTSQNVDDVDVLIPIFGESE-----DYDDLQGVVTGAVTSPVIDNI----- 1273  
QY 875 QNGLDDSHFISFEDISKTEGFRIRFINKETGNSIFIEKEIFSEYATHISKEISNIDK 934  
DB 1274 -----LSKIENEYEVLYKPLAG--VYRSLLKQLENNVMTF-----NVNVKD 1313  
QY 935 TI-----PDNV-NGKLVKKNLDAAEV-----NTLSAFPIQSILIENTYTKESI- 978  
DB 1314 ILNSRFNKNFNKNVLESDLPYKDLTSNYYVVDPYFLNKEKRDKFLSSYNYIKDSID 1373  
QY 979 SNLSVAMKVQVYAQLFSTGLNTITDASK 1006  
DB 1374 TDINFANDVLGYKILSEKYSGLDSIK 1401  
  
RESULT 10  
Y109 YEAST  
ID Y109 YEAST STANDARD; PRT; 1679 AA.  
AC P40457;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 195.1 kDa protein in DNA43-UBI1 intergenic region.  
GN Y1149C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
CC  
DR EMBL; M19143; AAA29653.1; --  
DR PIR; A54498; A54498.  
DR InterPro; IPR006209; EGF\_like.  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19  
FT CHAIN 20 1701  
FT CARBOHYD 110 110  
FT CARBOHYD 239 239  
FT CARBOHYD 470 470  
FT CARBOHYD 536 536  
FT CARBOHYD 607 607  
FT CARBOHYD 802 802  
FT CARBOHYD 899 899  
FT CARBOHYD 919 919  
FT CARBOHYD 965 965  
FT CARBOHYD 991 991  
FT CARBOHYD 1089 1089  
FT CARBOHYD 1196 1196  
FT CARBOHYD 1588 1588  
SQ SEQUENCE 1701 AA; 193719 MW; 3920875B73D38552 CRC64;  
  
Query Match  
Best Local Similarity 4.7%; Score 245; DB 1; Length 1701;  
Matches 213; Conservative 176; Mismatches 370; Indels 300; Gaps 51;  
  
QY 36 NMSESVVBEKYLKLDINNLTDNYLNTYKSGRNKALKKFKFVMEVLELKNLSLPVE 95  
DB 106 NFDNSDSTNTKYADLKHVQNYLFT-----IKELKYPEFLDNLNMLT-LS 152  
  
QY 96 KNLHFIWICQNDTRAINVQWQVNS--DYTVKFDVDSNAFLNLTAKTIVESATNLT 154  
DB 153 KNV-----DGFYLDGVEEINELLYKLNFFYD-----LLRAKLANDACANSYC 195  
  
QY 155 E-SFRENLDNPPFDYK-----FVRKMEIYKQKHFDYKSOIEENPEFIIDNIKT 209  
DB 196 QIPFNLIKIRANELDVLKLVFGYRKLPLDNKKNVGMKEDYIKN-----KTT 242  
  
QY 210 LSNEYSKOLEALNKYTEESLNKITANNNGDIRLEK--FADEDVRLYNQELVERNLAA 267  
DB 243 IAN-----INELIEGSKTIDQKNADNEEGKKLYQAVNYFLYKQLQEAHNLIS 294  
  
QY 268 ASDILRISLMDGGVLDVLDLPGIQLPKSI--NKPDSITNTSWEMKLEAIMKYKE 325  
DB 295 VLE-KRIDTLKKNENIKKLEIDIKITDAENPTTGSKENPLPENK-----KK 341  
  
QY 326 YIPGYTSK-----NFDMLDEVRQSFESALSSKSKSEIPLDLDIKVSPLEVKI 375  
DB 342 EVEGHEEKIKETAKTKFNIDSLFTD-PLEVYLRKKNKVDVTPKSDPTKSVQIPKV 400  
  
QY 376 AFANNSVINQALISLSDYCSLVINQIKNRY-KILNDNLNPSINSGTDFNTMTKIFSDK 434  
DB 401 PYPNGIVYPLPTDIHNSLAAD-----NDKNSYGLMNPDTKEKINE-----KIITD- 447  
  
QY 435 LASISNEDNMFMKITYLKVGPADPVRSTNLSGPGVYTGAYQDLLMFKONSTNIHL 494  
DB 448 -----NKRKIFI-----NNIKKQIDL----- 464  
  
QY 495 EPELRNFEPPKTKISOLTQETSLWSFNOARAKQFERYKGYREGALGEDNDLDAFON 554  
DB 465 -----EERKNINTKQNKKLE-----DYKSKDYELLEKFPYB-----MKPNN 505  
  
QY 555 TVLDDKYVSKKILSSMKTKNE---YIH-----YIVLQGDKISYEASCNLFSKDP 602  
DB 506 --FDKDVVD-KIFSARYTYNVEKQRYNNKFSNNNSVYVQKLKALSY-----LSD 554  
  
QY 603 YSSILYQKNIEGSETAYIVYV---ADAETIKDKIRIPYQISNKRN--IKLTFYG--HOK 655  
DB 555 YS---LRKGISEKDFNHYYTLKTGLEADIKKLE-----EIKSSENKILEKFNKGLTHSA 606

QY 656 SFENTDTFANLDV-DSLSEIETILNLAADISPKYIEINL-----LGCNMFPSYSIYAET 710  
DB 607 N-----ASLEVSIVKQVQKVLIIKKIEDLRK-IELFLKNAQLKOSIHVPNIYKPN 658  
  
QY 711 YPGK-----LLKIKDRVSELMPISIQ-----DSIT-----VSANQYEVIRINBEG--KR 752  
DB 659 KPPEYVLVLKKEVDKLEKFPKVKMDLKKQAVLSSITQPLVAASE-----TTDCGHST 714  
  
QY 753 EILDHSGKWINKEESIIKD-----ISSKEYISFNKPKENKIIVKSKYLHELSTL 800  
DB 715 HTLSQSGTEVTEETEVEETVGHVTTTITLPPKEESA--PKEVKVWENS----- 763  
  
QY 801 LOEIRNNANSSDIDLEKKVMLTECEINVASN-----IDRQIVEGRIEERAKNLTSDSI 852  
DB 764 ---IEHKSNDNSOALTKTIVLKKLDEFLTKSYCHKYLIVSNSSMDOKLLEVNLTPBEE 820  
  
QY 853 NYIKNEFKL-----IESISDSLYDLKHQGLDDSHFISPEDISKETENGPRIRINKETGN 907  
DB 821 NELKSCDPLDLLFNITQNNIPAMYSLYDSNMNIDLOHL--FFELYQKEMIYLLHKLKEENHI 878  
  
QY 908 SIFIFETEKEI-----FSEVATHISKEISNIKDTIFDNVNGKLVKK--VNL 950  
DB 879 KLLBEQKQITGTSSTSPGNTVNTAQSATHSNQOOSNASTNTONGVAVSSGPVV 938  
  
QY 951 DAAHEVNTLNSAFFIQSLIEVNTTKESLSNLSVAMKVQV 989  
DB 939 EESHDPILVLSI-----SNDLKGIVSLNLGNKTKV 969  
  
RESULT 12  
MSPI\_PLAF3  
ID MSPI\_PLAF3 STANDARD; PRT; 1682 AA.  
AC P19598; Q25921;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMWSA) (P190).  
GN MSP-1.  
OS Plasmodium falciparum (isolate ro-33 / Ghana).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5834;  
RN (1)  
RP SEQUENCE OF 1-1061 FROM N.A.  
RX MEDLINE=8816657; PubMed=3327688;  
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;  
"A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats."; EMBO J. 6:4137-4142(1987).  
RL (2)  
RP SEQUENCE OF 1032-1682 FROM N.A.  
RX MEDLINE=95354793; PubMed=7628566;  
RA Tolle R., Bujard H., Cooper J.A.;  
"Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1."; Exp. Parasitol. 81:47-54(1995).  
RL - SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).  
CC -!- KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; 235326; CAA84555.1; --  
 DR InterPro; IPR006209; EGF\_like.  
 DR Pfam; PF00008; EGF; 1.  
 KW Malaria; Merzoite; Polyporein; Repeat; Signal; Glycoprotein;  
 KW Transmembrane; GPI-anchor.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1682  
 FT TRANSMEM 1666 1682  
 FT CARBOHYD 233 233  
 FT CARBOHYD 462 462  
 FT CARBOHYD 528 528  
 FT CARBOHYD 599 599  
 FT CARBOHYD 785 785  
 FT CARBOHYD 881 881  
 FT CARBOHYD 901 901  
 FT CARBOHYD 947 947  
 FT CARBOHYD 1071 1071  
 FT CARBOHYD 1178 1178  
 FT CARBOHYD 1569 1569  
 FT CARBOHYD 1682 1682  
 SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;  
 Query Match 4.7%; Score 244; DB 1; Length 1682;  
 Best Local Similarity 19.8%; Pred. No. 0.02;  
 Matches 207; Conservative 180; Mismatches 374; Indels 284; Gaps 49;  
 QY 36 NMSSESVVEKYLKLDINLTNTYKSGRKNKALKKFKVLTWEVLEKNNSLTPVE 95  
 DB 100 NPSDDSSDAKADKRVQNYLT-----IKELKYPELFDLTHMLTCLD 147  
 QY 96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKPVYDSNAFLNFKTKTIVESATNNTL 154  
 DB 148 -NIH-----GFKLIDGYEIN-----ELLYKLN-FYFDLLRAKLVDCANDYC 189  
 QY 155 E-SFRENLDPEYDNK----FYKRMELIYKQKHFIYKYSQTEENPEFIIDNIITY 209  
 DB 190 QIPFNLKIRANELDVLKLVFGYRKPLDFIKONVGRMEDYIKKN-----KTT 236  
 QY 210 LSNEYSKOLEALNKYTEESLNTKNTANGNDIRNLEK---FADEDLRLYNQELVERWNL 266  
 DB 237 IAN-----INLEGGKTIIDQKNADNEGKKLVQAQVDLF-IYNKQLEAHNLI 287  
 QY 267 AASDILRLSMKEDGGVLDVILPGIQDLPKFSINKPDSITNTSWEMIKLEAIKMYKEY 326  
 DB 288 SVLE-KRIDTLKKNENIKLLEDIDKIDAEKPTTGVNQILS-----LRLEKESRHEBK 341  
 QY 327 IP--GWTSK-NFMDLDEEVQSFESALSSKSKSEIFLPLDLKVSPLVKIAPANNVI 383  
 DB 342 IREIAKTIKFNIDRLFTD-PLELEYLREKNKKVDVTPKSKQDPTKSVQIPKVPYNGIY 400  
 QY 384 NQALISLKDYSCLDVLINQIKRY-KILNDLNPISINEGTFDNTMTKIFSDKLASINED 442  
 DB 401 PLPLTDIHSNLAAD-----NDKNSYGLMMPHYTKKEKINE-----KIITD-----NKE 442  
 QY 443 NMFMKIKITNLYKVGPAVDRSTINLSGPGVYTGAQYQDILLMFKDSTNIHLLEPELRNPE 502  
 DB 443 RKIFI-----NNIKKQIDL-----EKNIN 462  
 QY 503 FPKTKISQTEQBITSLNFSNQARSKQFEYKKGVEGALGEDDNLDPAQNTVLDKQVY 562  
 DB 463 HTKEQNKLLLE-----DYKSKQYELLEKFEYEMFNKNFKD-----VVDKIFS 508  
 QY 563 SKKILSMKTR-NKEYIH-----YIVQLOGDKISYEASCNLSFKSDPYSSILYQKIEGSE 616  
 DB 509 ARYTNEVQRYNNKFSNNSSNVNVQKLKALSY-----LEDYS---LRKGLSEKD 557  
 QY 617 TAYYYV---ADAEIKEDKIRIPYQISKNRN---IKLTFIG--HGKSEFTDTFANLDV- 668  
 DB 558 FNNHYTLTKGLEADIKKLE-----EIKSENKILEKFNKGLTHSAN-----ASLEVS 605  
 QY 669 DLSLSEIEFILNAKADISPKYIEINL-----LGCNMFYSYIYAETYPCK-----LLLKIK 720  
 DB 606 DIVKLOVQVLLFKTKIEDLRK-IEFLKNAQLKDSIHVPNIYKPNQKPEFYLIYLKKEV 664

QY 721 DRVSELMF-----SISQDSITVSANOYE-----VRINEGKREILDHSGKW 761  
 DB 665 DKLEFIPKVKDMLKKEQAVLSSITQPLVAASETTEDGSHSTHTLSQSGETEVTETEET 724  
 QY 762 INKEESIIOKISSKEYISFNPKENKIIVSKYKYLHSLSTLLOEIRNANSSDIDLEKKVYL 821  
 DB 725 VGHITTVT-----ITLPPKEVKVENS-----IEHKSNDNSQALTKTVYL 764  
 QY 822 TECEINVASN-----IDRQIVEGRIEEAKNLTS-----DSINYIKNEFKLI 862  
 DB 765 KKLDFELTKSYCHIKYILVNSNMMDQKLEVVNLTPENELKSCDRLDLLFNQNNIPAM 824  
 QY 863 ESISDSL-YDLKHQGLDSDHSFISPEDISKTEGFRIFRINKETGNSIFIEKEI---- 917  
 DB 825 YSLYSMMNDLQH-----LFELYQKEMYYLHLKEENHIKKLEEPKQITGTS 874  
 QY 918 -----FSEYATHISKEISNIKTIFDNVNGKLVKK--VNLDAAHEVNTLNSAFF 964  
 DB 875 STSPGNTTNTVAQSAATHSNQSQNSNTQNGVAVSSGPAVVESHDPDLTVLSI-- 932  
 QY 965 IQSLIEYNTTKESLSNLSVAMKQV 989  
 DB 933 -----SNDLKGIVSLNLGNKTKV 951  
 RESULT 13  
 MSPI\_PLAFM  
 ID MSPI\_PLAFM STANDARD; PRT; 1701 AA.  
 AC P08569;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Merzoite surface protein 1 precursor (Merzoite surface antigens)  
 DE (PMSEA) (P190).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=70153;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88011243; PubMed=3079521;  
 RA Tanabe K., Mackay M., Goman M., Scaife J.G.;  
 RT "Allelic dimorphism in a surface antigen gene of the malaria parasite  
 Plasmodium falciparum".  
 RL J. Mol. Biol. 195:273-287(1987).  
 RN [2]  
 RP REVISIONS TO 1403; 1569 AND 1629.  
 RA Tanabe K.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-115 FROM N.A.  
 RX MEDLINE=86136024; PubMed=3004972;  
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
 Stunnenberg H., Bujard H.;  
 RT "Polymorphism of the precursor for the major surface antigens of  
 Plasmodium falciparum merzoites: studies at the genetic level".  
 RL EMBO J. 4:3823-3829(1985).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 (Potential).  
 CC -!- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
 MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X05624; CAA29112.1; --  
 DR InterPro; IPR006209; EGF\_like.

DR Pfam; PF00008; EGF, 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
Transmembrane; GPI-anchor.  
FT SIGNAL 1 19  
FT CHAIN 20 1701  
FT CARBOHYD 110 110  
FT CARBOHYD 239 239  
FT CARBOHYD 470 470  
FT CARBOHYD 536 536  
FT CARBOHYD 607 607  
FT CARBOHYD 802 802  
FT CARBOHYD 899 899  
FT CARBOHYD 919 919  
FT CARBOHYD 965 965  
FT CARBOHYD 991 991  
FT CARBOHYD 1089 1089  
FT CARBOHYD 1196 1196  
FT CARBOHYD 1588 1588  
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AP96EA98 CRC64;  
Query Match 4.7%; Score 244; DB 1; Length 1701;  
Best Local Similarity 20.6%; Pred No. 0.02;  
Matches 220; Conservative 174; Mismatches 357; Indels 316; Gaps 54;  
QY 36 NMSESVVEKYLKLDINNLTONLYTKSGRNKALKKFKYLTWELKNNSLTPVE 95  
DB 106 NPSNDSNSTKYADLKHVQNYLFT-----IKELKYPELFDLTNMLT-LS 152  
QY 96 KNLHFIWGGQINDFAINYNQWQVNS--DYTVKVFYDSNAFLNLTWKTIIVESATNTYL 154  
DB 153 KNV-----DGFYKYLIDYEEINELLYKLFYD-----LLRAKLNDACANSYC 195  
QY 155 E-SRENLDNPFQYK-----FYRKMEIYDKQHFIDYKKSQIBENPEFIIDNIKTY 209  
DB 196 QIPFLKTRANELDLVKLVGFRPLDNKKNVGMEDYIKN-----KIT 242  
QY 210 LSNEYSKDLALNKYIEESLANKITANNNGDIRNLEK--FADEBLVRLYNQELVERNLAA 267  
DB 243 IAN-----INELIEGSKTIDQKNADNEEGKKLYQAYNLFYINKQLQEAHNLIS 294  
QY 268 ASDILRISMLKEDGGVYLDVILPGIOPPLFSI--NKPDSTNTSWEMKLEAMKYKE 325  
DB 295 VLE-KRIDTLKKNENIKKLEIDIKTDAENPTTGSKENPLPNK-----KK 341  
QY 326 YIPGYTSK-----NFDMLDEVOVSFSSALSSKSDKSEIFLPDDIKVSPLEVKI 375  
DB 342 EYEGHEEKIKETAKTIKFNIDSLFTD-PLLEYLRKNKRVDTVPKSDPTKSYQIPKV 400  
QY 376 AFANNSVINQALISLSDYCSDLVINQIKNRY-KILNDNLNPSINEGTFDNTMTFSDK 434  
DB 401 PYPNGIVYPLPLTDIHNSLAAD-----NDKNSYGLDMLNPDTEKINE-----KIITD- 447  
QY 435 LASIGNEDMMFMKITYLKVGPADPVRSTNLSGPGVYTGAYQDLLMFKDNSTNIHLL 494  
DB 448 -----NKRKIPI-----NNIKKQIDL----- 464  
QY 495 EPELNFPEPKTKISLTQEBITSLWSFNQARAKSQFEYKKGYPGALGEDDNLDAQN 554  
DB 465 -----EKNINHTKQNKLLLE-----DYKSKDYELLEKPYE-----MKFNNN 505  
QY 555 TVLKDXYVSKILLSMKTANKE--YIH-----YIVQLQGOKISYEASCNLFSDKP 602  
DB 506 --FDKDVVD-KIFSAITYNVEKQRYNNKFSNNNSVNVNQLKXALSY-----LED 554  
QY 603 YSSILYQKNIEGETAYIVYV--ADAETIKEDKVRIPYQISKNR--IKLTFIG--HGK 655  
DB 555 YS--LRKGISSEKDFHNYTLTKGLEADIKKLT-----EIKSSENKILEKKNFKGLTHSA 606  
QY 656 SFENTDTFANLDV-DSLSEIETILNLAADISPKYIEINL---LGCNMFYSYSYABET 710  
DB 607 N-----ASLEVSQDIVLQVQKLVKIKIEDLRK-IEFLKNAQKOSIHVPNIYKPN 658  
QY 711 YPGK-----LLLKIKORVSELMPSISQ-----DSIT---VSAQYEVREINEEG--KR 752

DB 659 KPPEYLLIVLKKEVDKLEKFIKPKVKMDLKEQAVLSSITQPLVAASE-----TTEDGGHST 714  
QY 753 EILDHSGKWINKESIIND-----ISSKEYISFNPKNENKIIVKSKYLHELSTL 800  
DB 715 HTLSQSGETEVEETEVEETVGHVTTVTITLPPKBSA--PKVKVNVNS----- 763  
QY 801 LQEIERNANSDDIDLEKRVMLTECE-----INVA-SNIDRQIVEG---RIBEA 844  
DB 764 ---IEHKSNDNSQALTKTVYKLUDEPLTKSVYCHYKILVSNSSMDQKLLVYVNLTPEE 820  
QY 845 KNLTS-----DSINYIKNEFKLIESISDSL-YDLKHQNGLDSDHSPDISKTENGFRIR 899  
DB 821 KELKSCDPLDLLFNQNNIPAMISLYDSMNNDLQ-----LFFELYQKEMIIYVLH 870  
QY 900 FINKETGNSFIETEKET-----FSEYATHLSKEISNIKOTIDFNVNGKL 944  
DB 871 KLKEENHIKKLEEQKQITGTSSSTSPGNTVNTVAQSATHNSNQNSQNASSTNTQNGVA 930  
QY 945 VKK--VNLDAAHEVNTLNSAFFIQSLIEYNTTKESLNSLVAMKVQV 989  
DB 931 VSSGPVVEESHDPILTVLSI-----SNDLKGIVSLNLGNKTKV 969  
RESULT 14  
MSPI\_PLAFC  
ID MSPI\_PLAFC STANDARD; PRT; 1726 AA.  
AC P04934;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMSA) (P195).  
GN MSP-1.  
OS Plasmodium falciparum (isolate Camp / Malaysia).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5835;  
RN [1]  
RP SEQUENCE OF 1-1103 FROM N.A.  
RX MEDLINE=86205236; PubMed=3517809;  
RA Weber J.L., Leininger W.M., Lyon J.A.;  
RT "Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite Plasmodium falciparum.";  
RL Nucleic Acids Res. 14:3311-3323(1986).  
RN [2]  
RP SEQUENCE OF 1104-1726 FROM N.A.  
RX MEDLINE=88143999; PubMed=3278296;  
RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;  
RT "Merozoite surface protein sequence from the Camp strain of the human malaria parasite Plasmodium falciparum.";  
RL Nucleic Acids Res. 16:1206-1206(1988).  
CC -! SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).  
CC -! PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC EMBL; X03831; CAA27446.1; -.  
DR PIR; A23386; SAZQCM.  
DR InterPro; IPR006209; EGF\_like.  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
FT Transmembrane; GPI-anchor.  
FT SIGNAL 1 19  
FT CHAIN 20 1726  
MEROZOITE SURFACE PROTEIN 1.







FT	CARBOHYD	1221	1221	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1613	1613	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1726 AA;	196174 MW;	SB59CEEFA2F9A026 CRC64;
	Query Match	4.7%;	Score 243;	DB 1; Length 1726;
	Best Local Similarity	20.5%;	Pred. No. 0.023;	
	Matches 212;	Conservative 174;	Mismatches 359;	Indels 230; Gaps 52
QY	51	DINNLTDTNYLNTYKSGRNKALKFKPEYLWTMEVLKNNSLTPVEKNLHFIWIGSQINDT	110	: :     : :
DB	154	DLXHRVRYLFT-----IKELKYPELFDLTNHMLTLCD-NIH-----	189	: :     : :
QY	111	AINY-INQWDVNSDYTVKFVYDSNAFLINTLKKTIVESATNMLE-SFRENLNDPEFDY	168	: :     : :
DB	190	GFKYLIDGYEIN-----ELLYKLN-FYDFULRAKLANDVCANDCYQIPFNLIKIRANEUDV	243	: :     : :
QY	169	NK----PYRKMEIYDKQHFIIDYKSOIEENPEFIIDNIITKVLSNEYSKOLEALNKY	224	: :     : :
DB	244	LKKLVFGYRPLDNIKONVGKMEYIKYN-----KTYIAN-----INEL	282	: :     : :
QY	225	IIESLNKITANNGNDIRNLEK---FADEDLVRLYNQELVERWNLAASAADIIRISMLKBGD	281	: :     : :
DB	283	IEGSKTTIQKNADNEGKKLYQAQYDL--SIYNKQLAEAHNLISVLE-KRIDTLKQNE	340	: :     : :
QY	282	GVVLVDILPGIQ-PDLFKSINKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKPFMDLDE	340	: :     : :
DB	341	NIKELLDKINEIKNPFPANSNGTPTNTLDKNNKIIEHEE--KIRE-IAKTIKFNIDSFLT	397	: :     : :
QY	341	EVQRSPESALSSKDSSEIFLPLODIKVSPLVKIAFANNSVINQALISLSDSYCSDLVI	400	: :     : :
DB	398	D-PLELEYLRERKNKKVDVTPKSQDPKTSVOIQPKVPYPNGIYVPLPLTDIHNLSAAD---	453	: :     : :
QY	401	NOIKNRY-KITLNDMLNPSINEGTDFNTMTKIFSPDKLASISNEDNMFMKITYNLKVGFA	459	: :     : :
DB	454	-NDKNSYGLMNPTEKEINE-----KIITD-----NKRKIFI-----NNIK----	490	: :     : :
QY	460	PDRVSTINLSGPVYTGAQODLLMPKONSTNIHLLEPRLNPEFPKTKISOQTQEITSL	519	: :     : :
DB	491	-----KQIDLEEKKINHNTKEQNKKLL	511	: :     : :
QY	520	WSFNQARAQSOFEEYKKGYPEGALGEDDNLDPAQNTVLDKDQVSKILLSMKTRNKE---	576	: :     : :
DB	512	EDYE--KSKDDYELEKFBYB-----MKFNNN--FDKQVDV-KIFSARYTNVSKOR	558	: :     : :
QY	577	YIH-----YIVQLQGDKISIYEASCNLPSKDPYSSILYQKNIEGSTETAYVVV--A	624	: :     : :
DB	559	YNNKFPSSNNSNVNVQKLKALS-----LEDYS---LRKGISEKDFNHYYTLTKGL	607	: :     : :
QY	625	DAEIKEIDKYRIPIVOISNKRN--IKLTFIG-HGKSEFNTDTTFANLDV-DLSLSSIEITIL	679	: :     : :
DB	608	EADIKLTE-----EIKSSENKILEKNPKGLTHSAN-----ASLEVYDVIVKLQVQKVL	655	: :     : :
QY	680	NLAKADISPKEYIELNL-----LCGNMFSSIYAETYPKG-----LLLKKIDRVSELMP---	728	: :     : :
DB	656	LIRKIEDLRK-IIEUFLRNAQLKDSIHVENIYKPQNKPEFYILIVLKGEVDLKEFIPKVK	714	: :     : :
QY	729	-----SISQDSITVSANOYE-----VRINEEGREILDHSGWINKWEISIIKDI	772	: :     : :
DB	715	DMLKKEQAVLSSITOPLVAASETTEDGGHSHHTLSQSGETEVTETEET---TEEVV--GH	769	: :     : :
QY	773	SSKEYISFNPKENKIIIVKSKYLHELSTLLOBIRNNANSSDIDLEKKVMLMTECEINVASN-	831	: :     : :
DB	770	TTTTVTITLPPKEVKVENS-----IEHKNSDNSQALTCTVYVKKLDELPLTKSY	817	: :     : :
QY	832	-----IDROIUEGRIEEAANKLTSDSINYIK-----NEFKLIESISDSL-YD	871	: :     : :
DB	818	ICHKYILVSNSSMDQKLEVLNLTPEESENELKSQDPLDLLENFIQNNIIPAMTSLYDSMND	877	: :     : :
QY	872	LKHQNGLDSSHFIPEDISDKTENGFRPIRFINKETGNSIFITEKEI-----	917	: :     : :
DB	878	LOH-----LFFELYQEKMIYYLHLKLEENHIKKLBEOQKOITGTSSTSPGNTTV	927	: :     : :
QY	918	-FGEYATHISKEIGNIKDTIPDNVNGKLAVKK--VNLDAAHEVNTLNSAFFQSLIYEYNTT	974	: :     : :



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:26:27 ; Search time 30 Seconds  
(without alignments)  
8773.789 Million cell updates/sec

Title: US-09-126-816b-6\_COPY\_1\_1020

Perfect score: 5192

Sequence: 1 MNLVNAQLQKVMYVKFRIQ.....ITDASKVVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5171	99.6	2364	2 Q46342	Q46342 clostridium
2	4090.5	78.8	2367	2 Q9EXR0	Q9EXR0 clostridium
3	4081.5	78.6	2367	2 Q9F931	Q9F931 clostridium
4	4073.5	78.5	2367	2 Q46034	Q46034 clostridium
5	2144.5	41.3	554	2 Q93139	Q93139 clostridium
6	1911	36.8	698	2 Q86G53	Q86G53 clostridium
7	1893	36.5	698	2 Q9EXQ8	Q9EXQ8 clostridium
8	1887	36.3	697	2 Q86141	Q86141 clostridium
9	1469	28.3	553	2 Q9FCX5	Q9FCX5 clostridium
10	1424.5	27.4	2178	2 Q46149	Q46149 clostridium
11	411.5	7.9	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	408.5	7.9	3169	2 Q82916	Q82916 escherichia
13	386	7.4	3223	2 Q8L1L9	Q8L1L9 escherichia
14	384	7.4	3223	2 Q9RPH1	Q9RPH1 escherichia
15	381	7.3	3223	2 Q9RM48	Q9RM48 escherichia
16	381	7.3	3223	2 Q8VQR2	Q8VQR2 escherichia

17	377	7.3	3223	2 Q8VNR6	Q8VNR6 escherichia
18	362	7.0	3225	16 Q9PKM6	Q9PKM6 chlamydia m
19	357.5	6.9	2771	5 Q26216	Q26216 plasmodium
20	352	6.8	3317	16 Q8EWP8	Q8EWP8 mycoplasma
21	346.5	6.7	2752	5 Q8BJY0	Q8BJY0 plasmodium
22	345.5	6.7	1127	12 Q9YV76	Q9YV76 melanoplus
23	345	6.6	5767	5 Q81525	Q81525 plasmodium
24	338	6.5	2204	5 Q81L70	Q81L70 plasmodium
25	337.5	6.5	2747	5 Q9BJX9	Q9BJX9 plasmodium
26	335.5	6.5	2867	5 Q9N2M3	Q9N2M3 plasmodium
27	330	6.4	3724	5 Q77320	Q77320 plasmodium
28	324.5	6.2	1238	12 Q9EMP3	Q9EMP3 ansacta moo
29	318.5	6.1	2166	16 Q51465	Q51465 borrelia bu
30	316.5	6.1	3254	5 Q9BK45	Q9BK45 plasmodium
31	316.5	6.1	3255	16 Q9PKM8	Q9PKM8 chlamydia m
32	315.5	6.1	3130	5 Q81DX6	Q81DX6 plasmodium
33	315.5	6.1	3394	5 Q77384	Q77384 plasmodium
34	315	6.1	1453	8 Q9G9H3	Q9G9H3 schizophyll
35	315	6.1	3130	5 Q9BK46	Q9BK46 plasmodium
36	314.5	6.1	1387	5 Q9GZ76	Q9GZ76 plasmodium
37	313	6.0	3381	5 Q812V4	Q812V4 plasmodium
38	312	6.0	1956	5 Q81IE1	Q81IE1 plasmodium
39	312	6.0	5779	5 Q8IBS0	Q8IBS0 plasmodium
40	311	6.0	1302	2 Q49547	Q49547 mycoplasma
41	310.5	6.0	5687	5 Q81JH4	Q81JH4 plasmodium
42	310	6.0	2269	5 Q26223	Q26223 plasmodium
43	310	6.0	3335	16 Q9PKM7	Q9PKM7 chlamydia m
44	308	5.9	2792	5 Q814R2	Q814R2 plasmodium
45	305	5.9	1686	5 Q81FP9	Q81FP9 plasmodium

ALIGNMENTS

RESULT 1

Q46342	ID	Q46342	PRELIMINARY;	PRT; 2364 AA.
AC	Q46342;			
DT	01-NOV-1996	(TremBLrel. 01, Created)		
DT	01-NOV-1996	(TremBLrel. 01, Last sequence update)		
DT	01-MAR-2003	(TremBLrel. 23, Last annotation update)		
DE	Cytotoxin L.			
OS	Clostridium sordellii.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1505;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=6018;			
RX	MEDLINE=95369733; PubMed=7642137;			
RA	Green G.A., Schue V., Montell H.;			
RT	"Cloning and characterization of the cytotoxin L-encoding gene of			
RT	Clostridium sordellii: homology with Clostridium difficile cytotoxin			
RT	B.;"			
RL	Gene 161:57-61(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=6018;			
RX	MEDLINE=96149194; PubMed=8544213;			
RA	Green G.A., Schue V., Girardot R., Montell H.;			
RT	"Characterisation of an enterotoxin-negative, cytotoxin-positive			
RT	strain of Clostridium sordellii.;"			
RL	J. Med. Microbiol. 44:60-64(1996).			
DR	EMBL; X82638; CNA57959.1;			
DR	InterPro; IPR002479; CW_binding.			
DR	InterPro; IPR001950; TIF_SUII.			
DR	Pfam; PF01473; CW_binding_1; 18.			
DR	PROSITE; PS01118; SUII_1; 1.			
SQ	SEQUENCE 2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;			

Query Match 99.6%; Score 5171; DB 2; Length 2364;  
Best Local Similarity 99.7%; Pred. No. 4e-195;  
Matches 1017; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MNLVNAQLOKQVYVKFRIQDEYVAILNALBEYHNMSSESSVVEKYKLKLDINNTDNYL 60
DB 1 MNLVNAQLOKQVYVKFRIQDEYVAILNALBEYHNMSSESSVVEKYKLKLDINNTDNYL 60
QY 61 NTYKSGRNKALKKFKFXYLTMEVLELKNLSLTPVEKNLHFVIGGQINDTAINYNQWKD 120
DB 61 NTYKSGRNKALKKFKFXYLTMEVLELKNLSLTPVEKNLHFVIGGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTNLTSEFRENLDNDFPDYNNKFKYKMEIY 180
DB 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTNLTSEFRENLDNDFPDYNNKFKYKMEIY 180
QY 181 DKQKHFDYKSOIEENPEFIIDNIITKYLSEYKDLALNKYTEESLNKLTANNNGDI 240
DB 181 DKQKHFDYKSOIEENPEFIIDNIITKYLSEYKDLALNKYTEESLNKLTANNNGDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDPLFKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDPLFKS 300
QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEIF 360
DB 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEIF 360
QY 361 LPDLDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE 420
DB 361 LPDLDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE 420
QY 421 GTDFNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQD 480
DB 421 GTDFNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQD 480
QY 481 LLMFKDNSTNIHLLPELKNFPFPKTKISQLTEQETISLWSNQARAKSQFBEYKKGYPE 540
DB 481 LLMFKDNSTNIHLLPELKNFPFPKTKISQLTEQETISLWSNQARAKSQFBEYKKGYPE 540
QY 541 GALGEDNDLDPQNTVLDKDYVSKLSSMKRNKEYIHYIYVLOQDKLSYEASCNLFPSK 600
DB 541 GALGEDNDLDPQNTVLDKDYVSKLSSMKRNKEYIHYIYVLOQDKLSYEASCNLFPSK 600
QY 601 DYPSSILYQKNIEGSETAYYYVADAEIKEIDKYRIPYQISNKRNIKLTFFIGHGKSEFNT 660
DB 601 DYPSSILYQKNIEGSETAYYYVADAEIKEIDKYRIPYQISNKRNIKLTFFIGHGKSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCMNFSYSIVAETYPGKLLKTK 720
DB 661 DTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCMNFSYSIVAETYPGKLLKTK 720
QY 721 DRVSELMPISQDSITVSANOYEVRINEGKREILDHSGKWINKEESIHKDISKEYISF 780
DB 721 DRVSELMPISQDSITVSANOYEVRINEGKREILDHSGKWINKEESIHKDISKEYISF 780
QY 781 NPKENKIIVKSXYLHELSTLLOEIRNANSDDIDLEKKVMTLCEINVASNIDRQIVEGR 840
DB 781 NPKENKIIVKSXYLHELSTLLOEIRNANSDDIDLEKKVMTLCEINVASNIDRQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLSYDLKHQNGLDDSHFISPEDISKTENGFRIRF 900
DB 841 IBEAKNLTSDSINYIKNEFKLIESDSLSYDLKHQNGLDDSHFISPEDISKTENGFRIRF 900
QY 901 INKETGNSIFETEKEIFSEYATHISKEISNKTDTIFDNVNGKLVKKNLDAAEVNTLN 960
DB 901 INKETGNSIFETEKEIFSEYATHISKEISNKTDTIFDNVNGKLVKKNLDAAEVNTLN 960
QY 961 SAFFIOSLEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 SAFFIOSLEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020
```

RESULT 2  
Q9EXR0

ID Q9EXR0 PRELIMINARY; PRT; 2367 AA.

```
AC Q9EXR0;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Toxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RC STRAIN=8864;
RA von Eichel-Sreiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19891.1; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW binding 1; 17.
DR PROSITE; PS01118; SUI1.1; 1.
SQ SEQUENCE 2367 AA; 269072 MW; BF7087C92E7388E7 CRC64;

Query Match 78.8%; Score 4090.5; DB 2; Length 2367;
Best Local Similarity 77.2%; Pred. No. 1.1e-152;
Matches 788; Conservative 113; Mismatches 119; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKQVYVKFRIQDEYVAILNALBEYHNMSSESSVVEKYKLKLDINNTDNYL 60
DB 1 MSLVNRKQLEKXANVRVQDEYVAILDALSEYHNMSENTVVEKYKLKLDINSLTDYI 60
QY 61 NTYKSGRNKALKKFKFXYLTMEVLELKNLSLTPVEKNLHFVIGGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFXYLTVEILELKNLSLTPVEKNLHFVIGGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTNLTSEFRENLDNDFPDYNNKFKYKMEIY 180
DB 121 VNSDYTVKVFYDSNAFLINTLTKTIVESASNDTLESFRENLDNDFNHTAFPRKMQIY 180
QY 181 DKQKHFDYKSOIEENPEFIIDNIITKYLSEYKDLALNKYTEESLNKLTANNNGDI 240
DB 181 DKQKFINYKAAQKEENPDLIIDIVKTYLSNEYSKOIDEINAYTEESLNKVTENSGNDV 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDPLFKS 300
DB 241 RNFEFKTGEVFNLYEQELVERWNLGASDILRVAILKNIGVYLDVNLPGIHPDLFKD 300
QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVRSFESALSSKSKSEI 359
DB 301 INKPDVKTAVDWEEMQLEAIMKYKEYIPEVTSKHFTDLDEEVQSSFESVLASKSKSEI 360
QY 360 FLPLDDIKVSPLEVKIAFANNSVINQALISLKDSYCSDLVINQIKNRYKILNDNLNPSIN 419
DB 361 FLPLGGIEVSPLEVKVAFAGKSIIDQALISAKDSYCSDLIKQIQNRYKILNDTLGPIIS 420
QY 420 ECTDFTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYQ 479
DB 421 QGNDFTNTNFGESLGAINEENISFIAKTSYLRVGVFPEANTITLSGPTIYAGAYK 480
QY 480 DLLMFKDNSTNIHLLPELKNFPFPKTKISQLTEQETISLWSFNQARAKSQFBEYKKGYP 539
DB 481 DLLTFKEMSIDTSLSSLRNFPFPKVNISQATEQEKNSLWQFNEERAKIQFEEYKKNYF 540
QY 540 EGALGEDNDLDPQNTVLDKDYVSKLSSMKRNKEYIHYIYVLOQDKLSYEASCNLFPS 599
DB 541 EGALGEDNDLDPQNTVTDKEYLLEKISSTKSSERGYYVHYIVQLQDKISYEAAACNLFA 600
QY 600 KDPYSSILYQKNIEGSETAYYYVADAEIKEIDKYRIPYQISNKRNIKLTFFIGHGKSEFN 659
```

```

Db 601 KNPYDSILFQKNIEDSEVAYYNPTDSIEQIDKVRIPDRISDRPKIKLTLIGHGKAEFN 660
Qy 660 TDTFANLVDVSLSSSIEIETILAKADISPKYIEINLLGCNMFPSYIYAEETYPGKLLKI 719
Db 661 TDTFAGLDVDSLSSSIEIETILDADISPKSIEINLLGCNMFPSYVNVVEETYPGKLLLRV 720
Qy 720 KDVSELMPISQDSITVTSANQYEVRIINBEGKREILDHSGKWKINKEESIIDKISSKEYIS 779
Db 721 KDVSELMPISQDSITVTSANQYEVRIINBEGKREILDHSGKWKINKEESIIDKISSKEYIS 780
Qy 780 FNPENKIIVKSKYLHELSTLLQEIIRNNANSSDIIDLEKVKMLTECEINVASNIDRQIVVEG 839
Db 781 FNPENKIIVKSKYLPSTLLQEIIRNNANSSDIIDLEKVKMLTECEINVASNIDRQIVVEE 840
Qy 840 RIEEAKNLTSDSINYKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTENGFRIR 899
Db 841 RIEEAKNLTSDSINYKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTDEGFSIR 900
Qy 900 FINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKVKVLDAAHEVNTL 959
Db 901 FIDKGTGSIIFVETEKAFSEYANHITSEISKDTIPDTVNGKLVKVKVLDATHEVNTL 960
Qy 960 NSAPFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETI 1019
Db 961 NAAPFIQSLIGYNSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETI 1020
Qy 1020 D 1020
Db 1021 D 1021

```

RESULT 3

```

ID Q9F931 PRELIMINARY; PRT; 2367 AA.
AC Q9F931;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Cytocoxin B.
GN TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_TaxID=1496;
RP SEQUENCE FROM N.A.
RC STRAIN=5340;
RX MEDLINE=20448897; PubMed=10992443;
RA Sambol S.P., Merrigan M.M., Lyster D., Gerding D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
causes human clinical disease.";
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL; AF217292; AAG18011.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR001950; TIE_SUI1.
DR Pfam; PF01473; CW_binding_1; 18.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5EE715B5BD41E2F CRC64;

```

Query Match 78.6%; Score 4081.5; DB 2; Length 2367;  
 Best Local Similarity 76.8%; Pred. No. 2.4e-152; Indels 1; Gaps 1;  
 Matches 784; Conservative 116; Mismatches 120;

```

Qy 1 MNLVNAQIQKVVYKFRIOEYVAILNALAEYHNMSSESVVEKYKLDINLTDNYL 60
Db 1 MSLVNRKQLEKMANVRFRQDEYVAILDAALEHYHNMSSENTVVEKYKLDINLTDYI 60
Qy 61 NTYKSGRNKALKKPKYELTMEVLELKNLSLPVPEKNLHFIIWGGINDTAINYNQWKD 120
Db 61 NTYKSGRNKALKKPKYELTMEVLELKNLSLPVPEKNLHFIIWGGINDTAINYNQWKD 120
Qy 121 VNSDYTVKPYVDSNAFLINTLTKTIVESATNTLTESFRENLDNPPEDYNKFKYRKMELIY 180

```

```

Db 121 VNSDYTVKPYVDSNAFLINTLTKTIVESATNTLTESFRENLDNPPEDYNKFKYRKMELIY 180
Qy 181 DKQHFIDYKSKOLEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITTANGNDI 240
Db 181 DKQFNINYYKQAKKEENPDLIIDDIIVKTYLSNEYSKOIDEINAYIEESLNKVTENSNDV 240
Qy 241 RNLKFADEDILVRLYNQELVERWNLAAASDILRISMLKDEGCGVYLDVILPGIOPDLPKS 300
Db 241 RNPFEFKTGEVFNLYEQELVERWNLAGASDILRAILKNIIGVYLDVDMVLPCHPDLPKD 300
Qy 301 INKPSI--TNTSWEMIKLEALMKYKEYIPGYSKTFMDFMLDEEVQSPFSALSKSKDSBI 359
Db 301 INKPSVKTAVDWEEMQLEALMKYKEYIPGYSKTFMDFMLDEEVQSPFSALSKSKDSBI 360
Qy 360 FLPLDDIKVSPLEVKIAFANNSVINQALISLKDSCVSLVINQIKNRVKILNDNLNPSIN 419
Db 361 FLPLGDIIEVSPLEVKIAFAGSIINQALISLKDSCVSLVINQIKNRVKILNDNLNPSIN 420
Qy 420 ECTDFNTMTKIFSDKIASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQ 479
Db 421 QGNDENTMNFSGESLGAINEENISFIAKISYLRVGFPEANTTVTLSGPTIYAGAYK 480
Qy 480 DLLMFKDNTNHLLEPELRNPFEPKTKISQTEOEITSLMSFNQARAKSQPEEYKGYF 539
Db 481 DLLTFKMSIDTSLSSSELNPFEPKKNISQATEQEKNSLWQFNERRAKIQPEEYKKNYF 540
Qy 540 EGALGEDNDLPAQNTVLDKDYVSKLSSMKTRNKEYIHYIVOLQOGDKISYEASCNLPS 599
Db 541 EGALGEDNDLPAQNTVTDKEYLLEKISSSTKSSRGVYHYIVOLQOGDKISYEACNLPA 600
Qy 600 KDPYSSILYQKNIEGSETAYYVADABIKEDKVRIPYQISNKRNIKLTFTIGHGKSEFN 659
Db 601 KNPYDSILFQKNIEDSEVAYYNPTDSIEQIDKVRIPDRISDRPKIKLTLIGHGKAEFN 660
Qy 660 TDTFANLVDVSLSSSIEIETILAKADISPKYIEINLLGCNMFPSYIYAEETYPGKLLKI 719
Db 661 TDTFAGLDVDSLSSSIEIETILAKADISPKSIEINLLGCNMFPSYVNVVEETYPGKLLLRV 720
Qy 720 KDVSELMPISQDSITVTSANQYEVRIINBEGKREILDHSGKWKINKEESIIDKISSKEYIS 779
Db 721 KDVSELMPISQDSITVTSANQYEVRIINBEGKREILDHSGKWKINKEESIIDKISSKEYIS 780
Qy 780 FNPENKIIVKSKYLHELSTLLQEIIRNNANSSDIIDLEKVKMLTECEINVASNIDRQIVVEG 839
Db 781 FNPENKIIVKSKYLPSTLLQEIIRNNANSSDIIDLEKVKMLTECEINVASNIDRQIVVEE 840
Qy 840 RIEEAKNLTSDSINYKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTENGFRIR 899
Db 841 RIEEAKNLTSDSINYKNEFKLIESISDLYDLKHONGLDSDHSPFISFEDISKTDEGFSIR 900
Qy 900 FINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKVKVLDAAHEVNTL 959
Db 901 FINKETGNSIFVETEKTFSEYANHITSEISKDTIPDTVNGKLVKVKVLDATHEVNTL 960
Qy 960 NSAPFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETI 1019
Db 961 NAAPFIQSLIEYNSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETI 1020
Qy 1020 D 1020
Db 1021 D 1021

```

RESULT 4

```

Q46034
ID Q46034 PRELIMINARY; PRT; 2367 AA.
AC Q46034;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Toxin B.
OS Clostridium difficile.

```

OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC	Clostridium	
OX	NCBI_TaxID=1496;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RP	STRAIN=Isolate 1470;	
RX	MEDLINE=96079281; PubMed=7494480;	
RA	von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,	
RA	Sartingen S.;	
RT	"Closing in on the toxic domain through analysis of a variant	
RT	Clostridium difficile cytotoxin B.;"	
RL	Mol. Microbiol. 17:313-321(1995).	
DR	EMBL; Z23277; CAA80815.1; -	
DR	InterPro; IPR002479; CW_binding.	
DR	InterPro; IPR001950; TIF_SUI1.	
DR	Pfam; PF01473; CW_binding_1; 18.	
DR	PROSITE; PS01118; SUI1.1; 1.	
SQ	SEQUENCE 2367 AA; 269186 MW; EF9823DAE70427F3 CRC64;	
Query Match 78.5%; Score 4073.5; DB 2; Length 2367;		
Best Local Similarity 76.9%; Pred. No. 5e-152;		
Matches 785; Conservative 113; Mismatches 122; Indels 1; Gaps 1;		
QY	1 MNLVNAQLOKQVYKFRIOEDYVAILNALAEYHNMSSESVVEKYKLKDINNLTDNYL 60	
DB	1 MSLVNRKQLEKXANVFRVQEDYVAILDALAEYHNMSENTVVEKYKLKDINSLTDYI 60	
QY	61 NTYKSGRKNALKKPKFXYLTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120	
DB	61 DTYKSGRKNALKKPKFXYLTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120	
QY	121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTESFRENLDNDFYKFKRMELIY 180	
DB	121 VNSDYNVNVFDSNAFLINTLKTIVESATNTLTESFRENLDNDFYKFKRMELIY 180	
QY	181 DKQKHFIDYKYSQIENPEFIIDNIITKYLSEYKDLALNKYIEESLNKKTANNNGDI 240	
DB	181 DKQNFYNYKKAQKEENPDLIIDIVKYLSEYKDLALNKYIEESLNKKTANNNGDI 240	
QY	241 RNLKFADEDLVRLNQBELVERWNLAAASDILRLSMKDEGGVYLDVILPGIQDLPKFS 300	
DB	241 RNFEFKTGEVFNLYQEQLVERWNLAAASDILRLSMKDEGGVYLDVILPGIQDLPKFS 300	
QY	301 INKPDISI-TNTSEWIMTKALMKYKVIIPGYTSKNFMDLDEEVORSFESALSKSDKSI 359	
DB	301 INKPDISVTAVDEENQLEALMKHKEYIPEYTSKGFDTLDEEVORSFESALSKSDKSI 360	
QY	360 FLPLDDIKVSPLEVKIATFANNVINQALISLKDYSGLVINQIKNRYKILNDLNPSIN 419	
DB	361 FLPLGDIEVSPLEVKIATFAGKSIINQALISAKDSYCSLLIKQIKNRYKILNDLGLPIIS 420	
QY	420 EGTDFNTWKIPSDKLASISNEDNMFMKIITNLYKVGAPDVRSTINISGPGVYTAGYQ 479	
DB	421 QGNDFTTNNFESLGAIBANENISFIAKIGSYLRVGFYPEANTITLISGPTIYAGAYK 480	
QY	480 DILLMPKDNSTNHLPELARNFEFPKTKISQTEQITSLMSFNQARAKSQEYKGYF 539	
DB	481 DILLTFKMSIDTSLSSLELRFEPKKNISQATEQEKSLMQFNEERAKIQEYKKNYF 540	
QY	540 EGALGEDDNLDAQNTVLQDKYVSKKILSSMKTRNKEYIHYTVQLQDKISYEASCNLPS 599	
DB	541 EGALGEDDNLDSQNTVTDKEYILLEKISSKSGEGVHYIYVQLQDKISYEACNLPA 600	
QY	600 KDPYSILYKQKIEGSETAYYYVADAETIKDKYRIPYOIKNKNIKUTFTGHGKSEPN 659	
DB	601 KNPYDSILFQNIEDSEVAYYNTDSEIQEIDKYRIPDRISDRPKIKLFTFGHGAEPN 660	
QY	660 TDTFANLDVDSISSETITILNAKADISPKYILEINLLGCMFESYSYIAEETYPGLLKI 719	
DB	661 TDTFAGLDVDSUSSETIETAGLAKEDISPKSILEINLLGCMFESYSVNVETYPGLLLRV 720	
QY	720 KDRVSELMPSISQDSITVSANQYEVNRINEEGKREILDHSGKWINKKEESIIKDISKEYIS 779	
DB	721 KDRVSELMPSISQDSITVSANQYEVNRINEEGKREILDHSGKWINKKEESIIKDISKEYIS 780	
QY	780 FNPKNKIIVKSKYLHELSTLLOIRNNANSDDIDLEKVMLTECEINVASNIDRQIVEG 839	
DB	781 FNPKNKIIVKSKYKLPSTLLOIRNNANSDDIDLEKVMLTECEINVASNIDRQIVVEE 840	
QY	840 RIEEAKNLTSDSINYKNEFKLIESDSLYDLKHQGLDDSHFISPDISTKTENGFRIR 899	
DB	841 RIEEAKNLTSDSINYKNEFKLIESDSLYDLKHQGLDDSHFISPDISTKTENGFRIR 900	
QY	900 FINKETGNSFIETEKIEFSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAAEHVNTL 959	
DB	901 FINKETGESIFVETEKITFSEYANHITBEISKIGTIFDTVNGKLVKKVNLDTTDEVNTL 960	
QY	960 NSAFPIQSILIEYNTKESLSNLSVAMKVQVQAQLPSTGLNTITDASKVELVSTALDETI 1019	
DB	961 NAAFFIQSILIEYNTKESLSNLSVAMKVQVQAQLPSTGLNTITDASKVELVSTALDETI 1020	
QY	1020 D 1020	
DB	1021 D 1021	
RESULT 5		
Q93L39 PRELIMINARY; PRT; 554 AA.		
AC	Q93L39	
DT	01-DEC-2001 (TRENBLrel. 19, Created)	
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)	
DE	TcdB-C34 Cluster1-2 (Fragment).	
GN	TCD-B-C34.	
OS	Clostridium difficile.	
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC	Clostridium.	
OX	NCBI_TaxID=1496;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C34;	
RX	MEDLINE=21323121; PubMed=11430410;	
RA	Mehlig M., Moos M., Braun V., Kalt B., Mahony D.E.,	
RA	von Eichel-Streiber C.;	
RT	"Variant toxin B and a functional toxin A produced by Clostridium	
RT	difficile C34.;"	
RL	PEMS Microbiol. Lett. 198:171-176 (2001).	
DR	EMBL; AJ294944; CAC41640.1; -	
FT	NON TER 554	
SQ	SEQUENCE 554 AA; 63910 MW; 9285512AC1B1D87A CRC64;	
Query Match 41.3%; Score 2144.5; DB 2; Length 554;		
Best Local Similarity 73.3%; Pred. No. 5.5e-77;		
Matches 406; Conservative 67; Mismatches 80; Indels 1; Gaps 1;		
QY	1 MNLVNAQLOKQVYKFRIOEDYVAILNALAEYHNMSSESVVEKYKLKDINNLTDNYL 60	
DB	1 MSLVNRKQLEKXANVFRVQEDYVAILDALAEYHNMSENTVVEKYKLKDINSLTDYI 60	
QY	61 NTYKSGRKNALKKPKFXYLTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120	
DB	61 DTYKSGRKNALKKPKFXYLTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYNQWKD 120	
QY	121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTESFRENLDNDFYKFKRMELIY 180	
DB	121 VNSDYNVNVFDSNAFLINTLKTIVESATNTLTESFRENLDNDFYKFKRMELIY 180	
QY	181 DKQKHFIDYKYSQIENPEFIIDNIITKYLSEYKDLALNKYIEESLNKKTANNNGDI 240	
DB	181 DKQNFYNYKKAQKEENPDLIIDIVKYLSEYKDLALNKYIEESLNKKTANNNGDI 240	
QY	241 RNLKFADEDLVRLNQBELVERWNLAAASDILRLSMKDEGGVYLDVILPGIQDLPKFS 300	
DB	241 RNFEFKTGEVFNLYQEQLVERWNLAAASDILRLSMKDEGGVYLDVILPGIQDLPKFS 300	

```

QY 301 INKPSI-TNTSWMIKLBAMKYKEYIPGYTSKNFMDLDERVQSFESALSSKSKSEI 359
Db 301 INKPSVKTAVDEEWMQLEAIIMKYKEYIPEYTSKHPDTLDEEVQSFESVLASKNKSEI 360
QY 360 FLPLDDIKVSPLEVKIAPANNVINOALISLSDSYCSDLVINQIKRYKILNDNLNPSIN 419
Db 361 FLPLGDIKVPLEVKIAPAKGSIINOALISAKDSYCSDLIKQIQRYKTLNDTLGPIIS 420
QY 420 EGTDFNTWKIFSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQ 479
Db 421 QGNDFTNTWNGESLGAITANENTISPIAKISYLVGPGYPEANTITILSGTIYAGAYK 480
QY 480 DLLMFKONSTNTHLLEPELRFNFPFKTKISQTEQBITSLWFSNQARAKSQPEEYKGYF 539
Db 481 DLLTPKMSIDTSLSSLELRFNFPFKTKISQTEQEKSLWQFNEBRAKIQPEEYKGYF 540
QY 540 EGALGEDDNLDPAQ 553
Db 541 EGALGEDDNLDFSQ 554

RESULT 6
O68653 PRELIMINARY; PRT; 698 AA.
AC O68653;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 20309;
RX MEDLINE=99304497; PubMed=10376209;
RA Song K.P., Bai X.L., Chang S.Y.;
RT "Nucleotide and peptide sequences of the open reading frame encoding a
RT truncated toxin A gene of Clostridium difficile strain CCUG 20309."
RL DNA Seq. 10:93-96(1999).
DR EMBL; AF053400; AAC08437.1; -.
SQ SEQUENCE 698 AA; 80682 MW; 021BB268A3BDEC5E CRC64;

Query Match 36.8%; Score 1911; DB 2; Length 698;
Best Local Similarity 52.9%; Pred. No. 1.1e-67;
Matches 371; Conservative 136; Mismatches 186; Indels 8; Gaps 5;

QY 1 MNLVKAQIQKVMYVKFRIQDEYVAIINALBEYHNMSSESVVEKYKLKDINNLTNDYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKTVTLNDEYNKLTNNNNENKYLQKLKLNESIDVFM 59
QY 61 NTYKSGRNKALKKPKFKEYLTMEVLEKNLSLTPVEKNLHFPIWIGQINDTAINYNQWKD 120
Db 60 NKYKSSRRNALSNLKKDILKEVILKNSNTSPVEKNLHFVWIGGEVSDIVLEYIKQWAD 119
QY 121 VNSDYTVKVPYDSNAPLINTLTKTTIVESATNTNLESFRENLDNDFPDYKNFKRMEIY 180
Db 120 INAEYNIKLWYDSEAFVNTLKAIVESSTTALQLEEEIQNPQDFNNKFKYKRMFIY 179
QY 181 DQKQHFIDYKSOIEBENPFIIDNIIKTYSNEYSKDLEALNKYTESLNKITTANNNDI 240
Db 180 DQKQRFINYKQKQINPTVLTIDDIKSYLVSEYNRDETLLSEYRTNSLRKNSHGDV 239
QY 241 RNLEKFADELVRNLNQELVERWNLAAASDIIRISMLKEDGGVGLDVLDPGIDPLFKS 300
Db 240 RANSLFTEQELLNIYNQELNRCNLAAASDIVELLALKNFQGVGLDVLDPGLGHSDFKT 299
QY 301 INKPSITNTSWMIKLEAIMKYKEYIPGYTSKNFMDLDERVQSFESALSSKSKSEIF 360
Db 300 ISRPSIGLDRWEMIKLEAIMKYKEYINNTYENFDKLDQKDNPKLIIESKSEKSEIF 359
QY 361 LPLDDIKVSPLEVKIAPANNVINOALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420

```

```

Db 360 SKLENLAVSDLEIKIAPALGVSVINQALISKGSYLTNLVIEQIKNRYQFLNQHLPALPES 419
QY 421 GTDFNTWKIFSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480
Db 420 DNNFTDTTKIFHDSLFSNATAENSMLTKIAPYLVQGFEMPEARSTISLUSGPGAYASAYD 479
QY 481 LLMPKDNSTNTHLLEPELRFNFPFKTKISQTEQBITSLWFSNQARAKSQPEEYKGYPE 540
Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQTEQEIINSLSWSPDQASAKYQFEKYVKDYTG 539
QY 541 GALGEDDNLDPAQNTVLDKDY-VSKILSS--MKTRNKEYIHVIYVQLQGDKISYERASCM 597
Db 540 GSLSEDNEVDENKNTALDKYLLNKNKIPSNVVEEAGSKNYHYIIQLQGGDDISYEATCN 599
QY 598 FSKDPYSSILVQKNTGSETAVYYVAD--ARIKEIDKYRIPYQISNKNENIKLTPTIGHCK 655
Db 600 FSKNPNKSIITQRNM--NESAKSYFLSDGGSILSNKYRIPERLKNKEKVKVTFIGHCK 657
QY 656 SEFNTDTTFANLDDVDSLSSEIETILNLAKADISPKEYIEINLL 696
Db 658 DEFNTSEFARLSVDSLSSEISSFLDTIKLIDISPKNVEVNULL 698

RESULT 7
Q9EXQ8 PRELIMINARY; PRT; 698 AA.
AC Q9EXQ8;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE Truncated toxin A.
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19893.1; -.
SQ SEQUENCE 698 AA; 80799 MW; 66D9B21798D314BB CRC64;

Query Match 36.5%; Score 1893; DB 2; Length 698;
Best Local Similarity 52.4%; Pred. No. 5.4e-67;
Matches 367; Conservative 138; Mismatches 188; Indels 8; Gaps 5;

QY 1 MNLVKAQIQKVMYVKFRIQDEYVAIINALBEYHNMSSESVVEKYKLKDINNLTNDYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKTVTLNDEYNKLTNNNNENKYLQKLKLNESIDVFM 59
QY 61 NTYKSGRNKALKKPKFKEYLTMEVLEKNLSLTPVEKNLHFPIWIGQINDTAINYNQWKD 120
Db 60 NKYKSSRRNALSNLKKDILKEVILKNSNTSPVEKNLHFVWIGGEVSDIVLEYIKQWAD 119
QY 121 VNSDYTVKVPYDSNAPLINTLTKTTIVESATNTNLESFRENLDNDFPDYKNFKRMEIY 180
Db 120 INAEYNIKLWYDSEAFVNTLKAIVESSTTALQLEEEIQNPQDFNNKFKYKRMFIY 179
QY 181 DQKQHFIDYKSOIEBENPFIIDNIIKTYSNEYSKDLEALNKYTESLNKITTANNNDI 240
Db 180 DQKQRFINYKQKQINPTVLTIDDIKSYLVSEYNRDETLLSEYRTNSLRKNSHGDV 239
QY 241 RNLEKFADELVRNLNQELVERWNLAAASDIIRISMLKEDGGVGLDVLDPGIDPLFKS 300

```

Db 240 RANSLFTEQELLNIYNQELNLRGNLAASDIVRLLALKNFQGVYLDVMDLPGIHSDFLKT 299  
Qy 301 INKPSITWTSWEMIKLEALIMKYKEYIPGVTSGNFDMLDEEVQSPESALSCKSRIF 360  
Db 300 ISRPSSIGLDRWEMIKLEALIMKYKEYINNTYBENFDKLDQQLKDNFKLLIESKSEKEIF 359  
Qy 361 LPDIDDIKVSPLEVKIAFANNVINOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINE 420  
Db 360 SKLENLVSDLEIKIRFALGVSINOALISKGSYLTNLVIEQIKNRYQFNLQHLNPAIES 419  
Qy 421 GTDFNTMTKIFSKLASINEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480  
Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQVGFPEARSTISLSGPGAYASAYD 479  
Qy 481 LLMFKDNSTNIHLLEPELRNFPKTKISQLEQTEITSLSWFOARAKSQPEYKKGYPE 540  
Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQLEQTEINSLSWFOARAKSQPEYKKGYPE 539  
Qy 541 GALGEDNDLDFQNTVLDKDY-VSKKILSS--MKTRNKEYIHVIVOLQGDKISYEASCNL 597  
Db 540 GLSSENEVDFNKNTALDKNLLNKKIPSNVVEAGSKNYVHYIQLQDDDISYEATCNL 599  
Qy 598 FSKOPYSSILYQKNIEGSETAYYYVAD--AEIKEIDKYRIPYQISNKNRIKLTFTGHOK 655  
Db 600 FSKPNKNSIIQRM--NESAKSHFLSDGSEILELNKYRIPERLKNKEKVKVTFIGHCK 657  
Qy 656 SEFNTDTFANLVDLSLSSEIETILNLAADISPKYIEINLL 696  
Db 658 DEFNTSEFARLSVDSLSEISFLDTIKLDISPKNVENLL 698

## RESULT 8

086141 ID 086141 PRELIMINARY; PRT; 697 AA.  
AC 086141;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)  
DE TcdA protein (Fragment).  
GN TCDA.  
OS Clostridium difficile.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1496;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=8864;  
RA Soehn F., Leukel P., Weidmann M., Bichel-Streiber C.V., Braun V.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=8864;  
RA von Bichel-Streiber C.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y10689; CAA71690.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 697 AA; 80680 MW; CD650D3372D65B73 CRC64;

Query Match 36.3%; Score 1887; DB 2; Length 697;  
Best Local Similarity 52.4%; Pred. No. 9.3e-67;  
Matches 366; Conservative 137; Mismatches 188; Indels 8; Gaps 5;

Qy 3 LVNKAQLQKMYVYKFRIOEDYVAIINLAEEYHNMSESSVVEKYKLDKNNLTNYLT 62  
Db 2 LISKEELIKLAY-SIRPRENEYKTLTNLDEYNKLTNNNNENKYQLQKLNESIDVFMNK 60  
Qy 63 YKSGGNKALKKPKFVLTMEVLEKNNSLTPVEKNLHFWIGGQINDTAINYNQWKN 122  
Db 61 YKNSRRNALNKKKILKEVILIKNSGTSPEKNLHFWIGGEVSDIVLEYIKQWADIN 120  
Qy 123 SDYTVKFVDSNAFLNTLTKTTIVESATNNLTLESFRENLDNPDYKFKRMEIYDK 182  
Db 121 AYNIRLWDSFALVNTLKAIVESSTTEALQLLEEEIQNPQFDNMKFKRMEIYDR 180

Qy 183 QCHFDYYSQIOEENPEFIIIDIIKTYLSNEYSKOLEALNKYIEESLNKITANNNGDIRN 242  
Db 181 QKRFINYYSQINKPTVLTIDDIKSYLISEYNRDETILLESYRTNSLRKINSHGIDVRA 240  
Qy 243 LEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVDPGIDPLFKSIN 302  
Db 241 NSLFTQEELLNIYNQELNLRGNLAASDIVRLLALKNFQGVYLDVMDLPGIHSDFLKTIS 300  
Qy 303 KPDSTINTSWEMIKLEALIMKYKEYIPGVTSGNFDMLDEEVQSPESALSCKSEIFLP 362  
Db 301 RPSSTIGLDRWEMIKLEALIMKYKEYINNTYBENFDKLDQQLKDNFKLLIESKSEKEIFSK 360  
Qy 363 LDDIKVSPLEVKIAFANNVINOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINEGT 422  
Db 361 LENLVSDLEIKIRFALGVSINOALISKGSYLTNLVIEQIKNRYQFNLQHLNPAIESDN 420  
Qy 423 DFNFTMTKIFSKLASINEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYODLL 482  
Db 421 NFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQVGFPEARSTISLSGPGAYASAYDFI 480  
Qy 483 MFKONSTNIHLLEPELRNFPKTKISQLEQTEITSLSWFOARAKSQPEYKKGYPEGA 542  
Db 481 NLQENTIEKTLKASDLIEFKFPENNLSQLEQTEINSLSWFOARAKSQPEYKKGYPEGA 540  
Qy 543 LGEDDNDLDFQNTVLDKDY-VSKKILSS--MKTRNKEYIHVIVOLQGDKISYEASCNLFS 599  
Db 541 LSEDNEVDFNKNTALDKNLLNKKIPSNVVEAGSKNYVHYIQLQDDDISYEATCNLFS 600  
Qy 600 KOPYSSILYQKNIEGSETAYYYVAD--AEIKEIDKYRIPYQISNKNRIKLTFTIGHKSE 657  
Db 601 KPNKNSIIQRM--NESAKSHFLSDGSEILELNKYRIPERLKNKEKVKVTFIGHGKDE 658  
Qy 658 FNTDTFANLVDLSLSSEIETILNLAADISPKYIEINLL 696  
Db 659 FNTSEFARLSVDSLSEISFLDTIKLDISPKNVENLL 697

## RESULT 9

Q9FCX5 ID Q9FCX5 PRELIMINARY; PRT; 553 AA.  
AC Q9FCX5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Toxin A (Fragment).  
GN TCDA.  
OS Clostridium difficile.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1496;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C34;  
RX MEDLINE=20402122; PubMed=10931294;  
RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,  
RA von Bichel-Streiber C.;  
RT "A chimeric ribozyme in Clostridium difficile combines features of  
RT group I introns and insertion elements."  
RL Mol. Microbiol. 36:1447-1459(2000).  
DR EMBL; AJ131844; CAC03681.1; -.  
FT NON\_TER 553  
SQ SEQUENCE 553 AA; 63913 MW; DDE4551A6D8C3B25 CRC64;

Query Match 28.3%; Score 1469; DB 2; Length 553;  
Best Local Similarity 51.7%; Pred. No. 1.9e-50;  
Matches 287; Conservative 106; Mismatches 160; Indels 2; Gaps 2;

Qy 1 MNLVNAQLQKMYVYKFRIOEDYVAIINLAEEYHNMSESSVVEKYKLDKNNLTNYL 60  
Db 1 MSSIKKELIKLAY-SVRPRENEYKTLTNLDEYNKLTNNNNENKYQLQKLNESIDVFM 59  
Qy 61 NTYKSGRNKALKKPKFVLTMEVLEKNNSLTPVEKNLHFWIGGQINDTAINYNQWKN 120



```

Db 60 NRYKSSRRNALSNLKOTLKKEVILIKNSNTSPVEKNLHFVWIGVEVSDIALEYIKQWAD 119
Qy 121 VNSDYTKVPYDSNAPLINTLAKTIVESATNTNLESFRENLANDPEDYNNKFKYRKMEIY 180
Db 120 INAEYNKVLWYDSEAPLNTLKAIVESSTTALQLEEEIQNPQDM-KFYKQRMFIY 178
Qy 181 DKQKHFDIYKQIENPEFIIDNIKIYLSNEYSKDLALNKYIEESLNKITANNNDI 240
Db 179 DRQKRFINYKQINKPTVPTIDDIKSHLVSEYNRDETLLSEYRTNSLRKINSNHGDI 238
Qy 241 RNLEKPADEDLVRLYNOELVERWNLAAASDIIRISMLKEDGGVYLDVLDLPGIQDPLFKS 300
Db 239 RANSLFTEOELLNYSQELNLRGNLAAASDI VRLALKNFGVGYLDVLDLPGHSDLFKT 298
Qy 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNPDMLEEVQSPESALSSKSKSEIF 360
Db 299 IPRPSIGLDRWEMIKLEAIMKYKRYNNYTSNFDKLDQQKDNPKLIESKSEIF 358
Qy 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
Db 359 SKLENLWSDLEIKIAPALGVSINQALISKQSYLTNLVIEQKNRYQFLNQHLPAPES 418
Qy 421 GTDFNTTKIFSDKLASINEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480
Db 419 DNNFTDTTKIPHDSLFNSATAENSMLTKIAPYLQVGFMPPEARSTISLSGPGAYASAYD 478
Qy 481 LLMFQDNSTNIHLLEPELNFPFKTKISQLEQETSLWSNOARAKSQFEYKKGYPE 540
Db 479 FINLQNTIEKTLKASDLLEFPKNNSQLTEQELNSLWSPDQASAKYQIBRYVRYDYG 538
Qy 541 GALGEDDNLDPQNT 555
Db 539 GSPSGDNGVDNFKNT 553

RESULT 10
Q46149 PRELIMINARY; PRT; 2178 AA.
AC Q46149; Q46147; Q46148;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Alpha-toxin.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile."
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48636; CAA88565.1; -
DR EMBL; Z23280; CAA80818.1; -
DR EMBL; Z23281; CAA80819.1; -
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 10.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031C4A75A CRC64;

```

Query Match 27.4%; Score 1424.5; DB 2; Length 2178;  
 Best Local Similarity 32.7%; Pred. No. 4.8e-48;  
 Matches 342; Conservative 211; Mismatches 437; Indels 55; Gaps 19;

```

Qy 3 LVNKAQLOQMYYVVKFRIQEDYVAILNALALEEYH-NMSESSVVEYKYLKDKDINNLTNDYNL 61
Db 1 LITREQLMKIASIPLURKEPEYNLILDALENFRDIEGTSVKBYISKLSKLNELVDNYQT 61
Qy 62 TYKSGRNKALKKPEYLTMEVLEKNSLTP--VEKULHFIWIGQINDTAINYNQWK 119
Db 62 KYPSGGRNALENPDSLYSELRELTKNSRTSTIASKNLSFIWIGPISQDSLEYNNMWK 121
Qy 120 DVNSDYTKVPYDSNAPLINTLAKTIVESATNTNLESFRENLANDPEDYNNKFKYRKMEIY 179
Db 122 MENKYNIRLFDKNSLNLVNTLKTALIQESSKVIIEQNSNILDGTGYNKPYSDRMKLI 181
Qy 180 YDKQKHFDIYKQIENPEFIIDNIKIYLSNEYSKDLALNKYIEESLNKITANNNDI 239
Db 182 YRYKELKMWLENMKQNS--VDDIIINFLSNPKYDIGKLNKNKNNKNWIAIGATD 238
Qy 240 INLEKPADEDLVRLYNOELVERWNLAAASDIIRISMLKEDGGVYLDVLDLPGIQDPLFK 299
Db 239 I-NTENILTNLKSYYQELIQTNLAAASDIIRAILKKGVGVCYLDLDFPGVNLSPN 297
Qy 300 SINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNPDMLEEVQSPESALSSKSKSEI 359
Db 298 DISKPNCHDSNYWEAAIPEATANEKLMNYPKYMEQVPSEIKERILSPVRN-HDINDL 356
Qy 360 FLPLDDIKVSPLEV-----KIAFANNVINQALISLKDSYCSDLVINQIKRYKILNDNL 414
Db 357 ILPLGDIKISQLEILLRLKAATGKTFNAPFISNDSLTNLNLSQLENRYEILNSII 416
Qy 415 NP--SINEGTD--FNTTWKIFSDKLASINEDNMFMKIITNLYKVGFPADVRSTINLSG 470
Db 417 QEKFKICETYDSYINSVSELVLETPKNLSMDGSGFYOOIGYLSGSGFKPEYNSTVFPFG 476
Qy 471 PGVYTGAYQDLMPKDN-----STNIHLLEPELNFPFKTKISQLEQETSLWSFPQ 524
Db 477 PNIYSATCDTYHFINKTFDMLSSQOEIPEAS-NNLYFSK-----THDEFKSWLLRS 529
Qy 525 ARAKSQFEYKKGYPEGALGEDDNLDPQNTVLDKDYVSKKI--LSSMKTRNKEYIHYIV 582
Db 530 NIAEKEFQKLIKTYIGRTLNYEDGLNFKWKRVTTSSELLKVIIEVNSTKIYENYDLNML 589
Qy 593 QLOGDKISYEASCNLFSDKPYSSILYQKNIEG-SETATYVVYVADAEIKEDIKRYPYQIS 641
Db 590 QIQGDDISYESAVNVFGKNPKNSIL----IQGVDDFANVYFENGIVQSDNNINLSRPN 645
Qy 642 NKRNIKLTFIGHGKSEPTDTFANLDDVDSLSEI-----ETILNLAKADISPKYTEINIL 696
Db 646 DIKKIKLTLIGHGENVFNPKLPFGKTVNDLYTNIIKPKQLHLEREGVILKNKYLKINIL 705
Qy 697 GCNMFYSIYABETYPGKLLKIKDRVSELMP-SISQDSITVSANQYEVRIINEEGKREIL 755
Db 706 GCMFTPKVDINSTFVGKLFNKIS--RDLPKGFSGKQLEISANKYAIRINREGKREVL 762
Qy 756 DHSGKWNKERSIINKDISKEYISNPKNKIIVKSKYLHELSTLLOEIRNNANSSDIDL 815
Db 763 DYFGKWSNTDLIAEQISNKYVYVWNEVENTLSARVEQLNKVAFAKDI-----NSIIQT 817
Qy 816 EKKVMTCEINVASNIDRQIVEGRIERAKNLTSDSI---NYIKNEFKLIESISDSLYD 871
Db 818 TNNQELKQSLVNTYADLITLYSELKEDIPELDNIQIKERIIINEISRLHDFNSIILD 877
Qy 872 LKHQNGLDDSHFISFEDISKTENGFRIRPINKETGNSIPFETEKEIFSFYATHISKESIN 931
Db 878 FYQKNISNNMIILFDSIIKEKDYNNVKLANKITGETSVIKTYSDSLWNFTNKKYKIVDD 937
Qy 932 IKDITFDNVNGLVKKNLNDAAHEVNTLNSAPFIOSLIEYNTTKESLSNLSVAMKVQVTA 991
Db 938 IKGIIVKQINGEPIKADFEBQNPSSLNSAMLQMLLDYKPYTEILTNNMTSLKVQAYA 997
Qy 992 QLFSTGLNTITDASKVVELVSTALD 1016
Db 998 QIFQLSIGAIQAEITVITIISDALN 1022

```

```
RESULT 11
Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
AC Q9ZGR4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative cytotoxin.
GN L7095.
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
[1] SEQUENCE FROM N.A.
RP STRAIN=EDL933;
RC MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
of Escherichia coli O157:H7."
RL Nucleic Acids Res. 26:4196-4204 (1998).
DR EMBL; AF074613; AAC70163.1; -.
DR InterPro; IPR006473; Yop.T.
DR TIGRFA; TIGR01586; yopT_cys_prot; 1.
KW Plasmid.
SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BD5FD1 CRC64;

Query Match 7.9%; Score 411.5; DB 2; Length 3169;
Best Local Similarity 20.9%; Pred. No. 4.3e-08;
Matches 257; Conservative 195; Mismatches 453; Indels 323; Gaps 54;

QY 14 YKPRI---QBEYVAILNAL--EYHNMSSESVVEKYLKDINLNTYNTYKSKR 68
DB 160 YIKIRTRGAEDQTTITQSLIINELLNGVDRNTI--PFQKISELNDIHSYENMQIKSR 218
QY 69 N--KALKKPKVYLTMEVLEKNS-----LTPVEKNLHFTWIGQ 106
DB 219 KGIEILVKGQELSSLINONKGNKQSDNASKIINLLGIEYQSHKVDIEPFIHAWVAGA 278
QY 107 INDITAINYNQKDVNSDYTVKPYDSNAP-----LINTLKT----- 144
DB 279 PDNTSYITAFNTKYDYTYLLWIDPNAFGAAPSGILKNIAMVYAIMLRRTNPHLAE 338
QY 145 -----IVESATNTLE--SPRENLDPEFY-----NKFYKRMIEIYDQKHFI 187
DB 339 EMNEVILKIQNETIEFKETRELRKELNRYKSLTSETKGFNVFFLESMIGMQDNFY 398
QY 188 DYVKQIENPEFI-----IDNLIK--TVLSNEYKDLALNKYIEESLNKITANNGN-- 238
DB 399 TYCISNGISNTDDISLDFLTNVKLKSPVQNDKSTVEKNRKRDIIDLKNTISQKPGDRF 458
QY 239 ---DIRNLEKFADEDLVRLYNQBLVERWNLAAASDILRISMLKEDGGVYLDVILPGIQP 295
DB 459 QLARDINTLESFKPKPDQYFFQEMLLRWYAAASDQVRINILKEYGIYTDITDILPAYSD 518
QY 296 DLPKSNKPDSTNTSWEMIKLEAIMKY--EVIPO--YTSKNFMDLDEVOVSFESALS 351
DB 519 KVSQIINE--KSDDKRFFEDLKLRIISSESILSLIKGEYSIKH--DGLDETTLNQLNML- 575
QY 352 SKSDKSEIFLPLDDIKVSPLEVKIAPANNSVI-----NOALISLKDYSYC 395
DB 576 SEIEK-----LTIDY--FKPVETKVRDTPKIFRYQKWTENTWIRGNMFMTHKGSKC 630
QY 396 SDLVINQIKNRYKILNDLNINPSINEGTFDNTTWMKIFSDKLASTSNE-----DNM 444
DB 631 IDFILSQKKQYLEL--QRIRDNISYNNLFYTT-----EDLKSINNVAGIIPAKKYLEHG 684
QY 445 MFMIKITYLKVGAFDPVSTINLSGP----- 471
DB 685 LF-----SEYRQDGTPIYVSTLNSISGPDIMIQMKYKYSGLRIGEVHIDKNKLSVDNPL 740
```

```
QY 472 GVTGAYQDLLMF---KDNSTNIHLLEPE-----LRNFEPKTKISOLT-----EQEITS 518
DB 741 GVIASSNKDNKSNFWLNPVSVGINDIITPDDSSWAVRNNDINKILFEKINCHVPEKLPFS 800
QY 519 LWSFNQARAKSQFBEYKGYGEGALGEDDDNDFQNTVLDKDYVSKKLLSSMKTRNKYI 578
DB 801 LY-----YEIDSRSPFGW---DNKSIKHVTEINKDLI--KDINLLLTSSNDIV 844
QY 579 HYIVQLQDKISYEASCNLFPSKDPYSSILYQKNIEGSETAYVYVYVADAEIKIDKRYPI 638
DB 845 KLLIKL--DRELYAISKI-----DNPLALRSIRTLOQLANYVYVNTFPENTINIIY 896
QY 639 QISNKRN-----IKLTFIGHKSEFNTDTFANLDVDS-----LSSEIET 677
DB 897 DFYRKQDDLLSAIKL-----FSRNDADTKIIVWYNSVMKVNFLREVISC 942
QY 678 ILNLAKADISPKYIEINLLGCMFYSIYAEYTPGKLLKIKORVSELMPISISODSI-- 735
DB 943 VLRSKKVD---SYINEN-----KKNLKSKEDA--GALRDYAKLKKMELFSLMLDDGGYK 990
QY 736 TVSAQYEVREINEEGKREILDHSGKWINKESIIDISSKEYISFNP-----K 783
DB 991 IITNAY---IKERDKL-----SGIYNIENSIHGESFDIIRSNQHEWGLSTVQPK 1042
QY 784 ENKIIVKSKYLHELSTLQEIIRNNANSSDIDLEKKVMLTECEINVAS-----NIDQIVE 838
DB 1043 KFEFVVKSE--LSSAKSIFDDIKNK--YITDPETKRNVLVHQLSDDIKERIAFLDISHYAVP 1100
QY 839 GRIEAKNLTSDSINYIKNEPKLIESI-----SDSLYDLKHQNL 878
DB 1101 GSLLEKQLS---GYVFSINIIAEYLLASVSGHYSHGVVPAPSDKLELLRRHTK 1156
QY 879 DDSHFISFISDKTENGFRIRFINKETGNSI---PIETEKEIFSEYATHISKEISNIKDT 935
DB 1157 SNSEMI--EKITP-----YVYDILSNVSNVRPPLSEEQKILNDIKLEISKVS--EQ 1207
QY 936 IPDNVNGKLVKKVNLDAAEVNTLNSAFFIOSLIEYNTTK-----ESLSNLSV----- 983
DB 1208 YFMKTEQKSSVIGIKYSVDPRYNEINFLSLPINQNLTLPMYRYFEMLYDIHIGIEN 1267
QY 984 -AMKVQVYAQLFSTGLNTITDASKVVEL 1010
DB 1268 KANREFIYKFSLSNLDFLINDERVLNL 1295

RESULT 12
O82916 PRELIMINARY; PRT; 3169 AA.
AC O82916;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Toxin B.
GN TOXB.
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
[1] SEQUENCE FROM N.A.
RP STRAIN=O157:H7;
RC MEDLINE=98290540; PubMed=9628576;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamauchi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
outbreak."
RL DNA Res. 5:1-9 (1998).
DR EMBL; AB011549; BAA31815.1; -.
DR InterPro; IPR006473; Yop.T.
```

DR TIGR01586; yopT\_cys\_prot; 1.

KW Plasmid.  
SQ SEQUENCE 3169 AA; 361943 MW; D9BED726A62DDBFF CRC64;

Query Match 7.9%; Score 408.5; DB 2; Length 3169;  
Best Local Similarity 20.7%; Pred. No. 5.7e-08;  
Matches 255; Conservative 200; Mismatches 449; Indels 325; Gaps 55;

QY 14 YVKFRI--QDEYVAINAL--EEYHNMSSESVVEKYLKADINNLTDNYLNTYKKSQR 68  
DB 160 YIKIRTRGAEDOTTITQSLLINELLNGVDRTI--PFQISELANDIHSYENMQIKNSR 218  
QY 69 -----NKKAKPKPE-----YLTMEVLEKNSLTPVEKNLHFIWIG 105  
DB 219 KGIEILVKGELLSSLINVKNKQLSDNASKIINLLGIEYQSHKV-DIEPPIHAWVAG 277  
QY 106 QINDTAINVQWQVNSDYTKFYVDSNAP-----LINTLTKT----- 144  
DB 278 APPDNTFSYITAFNTYKDYTYLLWIDPNAGAAKFGSGILKNIAMNYAIMRLRRTNPHLA 337  
QY 145 -----IVESATNTLE--SFRENLDPEFY-----NKFVRKMEIYDKOKHF 186  
DB 338 EMEVEILKIQINQETIEFKETRELKELNRYKSLTSEKKEFNVFLESNIGMQNY 397  
QY 187 IDYKQSIENPEFI-----IDNIK--TYLSNEYSKOLEALNKYIBESLNKITANNNG- 238  
DB 398 FTYCISNGISNTDDISRLDFTNLVLSPEVQNDPKSTVEKNRDRIDLLKNTISQKFGDR 457  
QY 239 ----DIRNLEKFADEDLVRLNQELVERWNLAASDILRISMLKEDGGYLDVLDLPGIO 294  
DB 458 FOLRDINTLESFKPQDYFFYQOEMLRWNYAASDQVRINTLKEGYGYITDITLPAYS 517  
QY 295 PDLFKSINKPDSITNTSEMIKLEAIMKYK--EYIPG--YTSKNFMDLDEEVOVSFESAL 350  
DB 518 DKVQSIIIE--KSDRRKFFEDLKLRIISSEIUSLNGEKYSIKH--DGLDETTLNQLNWL 575  
QY 351 SKSKDKSEIFLDDIKVSPLEVKIAFANNSVI-----NQALISLKDSY 394  
DB 576 -SEIEK-----LTIDY--PKPVETKVVRDTFKPKYQKWNTWIRGNMPEMLTHKSK 629  
QY 395 CSDLVINQIKRYKILNDNLNLSINEGTFNTMKIFSCLKASINE-----DN 443  
DB 630 CIDFILSGQKQYLE-QRIRDNISYNNLFYTT-----EDLSLNNVAIGGIPAKKYLEH 683  
QY 444 MMFMIKITNLYKGPAPVRSSTINLSGP-----NKKYKYSIGRIGEVHKNKLSDNVF 471  
DB 684 GLF-----SEYRGDTIPYVSTINLSGPDIMQMKYKYSIGRIGEVHKNKLSDNVF 739  
QY 472 -GVYTGAYQDLMP---KDNSTNIHLLEPE-----LRNFEPKTKISOLT---EQEIT 517  
DB 740 LGVYASSNKDNKSFNWLNPVSVGINDITPDDESSWVRNDINKILFEKINCHVPEKLT 799  
QY 518 SLWSFNQARAKSQPEYKGYEGALGEDNDLDPQNTVLDKDYYSKILSSMKTRNKBY 577  
DB 800 SLY-----YEIDSRSPFGW---DNKSIXHVTEINKDLI--KQINLLTSSNID 843  
QY 578 IHVIVQLQDKISYEASCNLFPSKDPYSSILYQKNLEGSETAYYYVADAEIKDYRIP 637  
DB 844 VKLLIKL--DRELYAISKI-----DNPLAURSIRTLQLQANVYTSNTPFENTINFI 895  
QY 638 YQISNKRN-----IKLTFIGHKSEFNTDTPANLDVDS-----LSSEIE 676  
DB 896 YDFYRKQDALLSAKL-----FSRNDADTKIIVWYNSVMEKNVFLREVLS 941  
QY 677 TILNLAKADISPKYIEINLLGNMPSYSIYABETVPGKLLKIKORVSELMPSISQDSI- 735  
DB 942 CVLRSKVD--SYINEN-----KNLSKEDA--GALRDYAKLKMELFSMLDDGDKY 989  
QY 736 -TVSANQVEVRINEEKGREILDHSGKWINKESIIKDISKEYISFNP----- 782  
DB 990 KIITNAY---IKERDKL-----SGIIVNIENSIIISGHESFDIIRSNQHEWGDLTVRQF 1041  
QY 783 KENKIVSKYLHELSTLIQEIIRNNANGSDIDLEKKVWLTECEINVAS-----NIDRQIV 837

DB 1042 KKFEPYVXSE--LSSAKSIFDDIKNK-YITDPETKRNVLHYQLDSIKERIAFLDSHVAY 1099  
QY 838 EGRIEEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNG 877  
DB 1100 PGSLEKIKQLS-----GYVFSINIILAEYLLASVGVSHGVVVPAPSDKLELURRHT 1155  
QY 878 LDDSHFISFEDISKTENGFRIFINKETGNSI---FIETEKEIFSEYATHISKEISINIKD 934  
DB 1156 KSNSEMI--EKITP-----YVVDILSDNVSNVLRPPLSEQKKILNDIKLEISKSVS---E 1206  
QY 935 TIFDNVNGKLVKKNVLDAAHEVNTLNSAFFTQSLEIYNTTK-----ESLSNLSV--- 983  
DB 1207 QYFMKLTQKSSVIGIKYSVDFDRYENLFLSLPINQNLTLPFMYRYFEMLYDIHIGITE 1266  
QY 984 --AMKVQVYAOQLFSTGLNTITDASKVVEL 1010  
DB 1267 NKANREFIYSKFPSSLNLDLFLINDERVLNL 1295  
RESULT 13  
Q8L1L9 PRELIMINARY; PRT; 3223 AA.  
ID Q8L1L9;  
AC Q8L1L9;  
DT 01-OCT-2002 (TremBLrel. 22, Created)  
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE EfaI-Lymphostatin-like protein.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=493/89;  
RA Janka A.;  
RT "Identification and Distribution of the Enterohemorrhagic Escherichia coli Factor for Adherence (efal) Gene in Sorbitol Fermenting Escherichia coli O157:H-";  
RL Thesis (2002). Department of Institut fuer Hygiene, University of Muenster, Muenster, Germany.  
DR EMBL; AJ459584; CAD30848.1; -  
DR InterPro; IPR001917; Ntrtransf.2.  
DR InterPro; IPR00169; SHprot.acsite.  
DR TIGR01586; yopT\_cys\_prot; 1.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 1.  
SQ SEQUENCE 3223 AA; 365885 MW; 1A65A3B0A7C6B6ED CRC64;  
Query Match 7.4%; Score 386; DB 2; Length 3223;  
Best Local Similarity 19.9%; Pred. No. 4.4e-07;  
Matches 255; Conservative 203; Mismatches 433; Indels 392; Gaps 55;  
QY 2 NLVNAQAQKQVYKVFRIQDEYVAINALBEYHNMSSESS-----VVEKYLK-LKDINN 54  
DB 226 NFVKISLSDELTKY---ANEILLEIKMEYNLLPKNSRNGKLKQADLKLIME 282  
QY 55 LTDNYLNTYKSGRKNALKPKFELYTMVELEKN---NSLTPVEKNLHFIWIGQINDTA 111  
DB 283 DTSVTENTFKN-----IEMAADIKREYYSHTVDIEKNIHAIWAGSPESI 329  
QY 112 INYINQWQVNSDYTKFYVDSNAPLINTLTKTIVESATNTNLTESFRENLDPEFYK 171  
DB 330 SDYIKTFLTKYKFTPYLWVDEKAFGAAKFTSVLKQIAPDLACRTIQONTPOKNIDFINL 389  
QY 172 Y---RKR-----MEIYDKQKHFIY--- 189  
DB 390 YNEIRKYNPNPSGOEYLNKRLRYATYQKISTPLKHEFNSFPLENMKLQDNFNYCI 449  
QY 190 YKQSIENPEFIID---NIITYLSNE-----YSKDLKALNKYIBESL-----NKT 233  
DB 450 VKGVTINDELRLYNLKNVVK--LSDDDIGNYQKTINDKORVKKLILDLQKQFGENRIS 507

```
QY 234 ANNGDIRLEKFADEDLVRLYNQELVERWNLAAASDILRLSMKEDGGVYLDVILPCI 293
DB 508 IKDVNSLTSLSKSENN---HNYQTMLLRWNPYPAASDILRMVILKEHGGIYTDMPAY 564
QY 294 QPDLFINKPSDITNTSWEMIKL-----EAMKYKEYIPGYTSKNFMDLD--EEVQSPFE 347
DB 565 SKQVIFKIMQMGNDNRFLDLKLRRAISDGLVRY-----VNNQNIDEVVNEYI----- 613
QY 348 SALSSKSKSEIFLPDLDIKVSPLEVKIAPANNSV-----INQ 385
DB 614 ----SDADKNIKKILTEISKMPEDSIFTKINTRIPROTMPIRLRYHLWPGWNRGLNG 669
QY 386 ALISLKDSYCSDLVI---NQIKNRYKILNDNLNPSINEGTDFNTMTKIFSDKLASINED 442
DB 670 FMLSHKSGEVIDAVIAGQNAVRELRIRDNHSEIY---FKQT-----DELSLDPD 720
QY 443 NMFMIKITNYLKVG-----FAPDVRSTINLSGP----- 471
DB 721 KIGGIL-VKKYLSGSLFSKFRQDTIPEALSTLQISGPDLIQKMLQFRSRGVLEBEFI 779
QY 472 ----GVY-----TGAYQ-----DILMFKN 488
DB 780 NERKLSKAYIGYKTTGTCKYDMLTPESIGVNDVTPADESTWCIGKRCVDDFLFKQVS 839
QY 489 TNLHLEPELRNPFPPKTKISQLTEOITSLWSFNQAR-AKQSPFEYKKGVEGALGEDD 547
DB 840 T-----LKTENLPFLTKIDTDTFFSQNSTTKKQDKQKIDQTLVRYNE--LIDSS 889
QY 548 NLDFQAONTVLDK-----DYVSKTILSSMKTRNKEYIHVIVQLQDKTISYASCNL 597
DB 890 TIDFKNLYEIDQMLHIMLEMNDIAKRSLSFLOVQIAEKIRMT-----IPVDNINI 943
QY 598 FSKDPYSSILYQKNIEGSTAYYYYVADAEIKEDKIRIPYQISNKNRNIKLTFIGHKSE 657
DB 944 YPD-----LHKNDNDLSMKGFLAS--NPHTKINILYSNKTEHNI----- 983
QY 658 FNTDTPANLDVLSSEIETILNLAKADISPKYIEINILGNCNPSIYAEETYPCKUL- 716
DB 984 FIKOLFS---FAMVENLREDIINNMSKDKTP-----ENWEGRVML 1020
QY 717 ----LKIKDRVSELMPSISODSITVSANQYE---VRINEEG-KREILDHS--GKWINK 765
DB 1021 QRYLELKMKHLSLQSSQEAPELEISTFIYENDFLREKIEAVKKNQNSHLYFEKIKKE 1080
QY 766 ESIIDKISKEY-----ISFNPKNKIIVSKYLHELSTLLQERINNANSSDIDL 815
DB 1081 QNTQDLSTKEQKQLIKALKKEISGNTKED---SHYDRLLDAFLK--KHNE--I 1128
QY 816 EKKVMLTECEINVAS-----NIDROIVEG---RIEAKNLTS- -SINY 854
DB 1129 HNKIQIKDFEKEYSVAIHNDKIVFGQTDRLRYHEGVFSDINTLRYTLHLGIGTG 1188
QY 855 IKNEFKLIESISDLYDLKHONGLDSDHSFISPEDISKTENGFRIRPINKETGNSPIETE 914
DB 1189 VHTENLLPAPSSLLINILKE-----HYNEDEISAKPLAYDY-ILNKESSESPVEIL 1241
QY 915 KEIFS-----EYATHISKETSNIKDTIFDNVNGKLVKV----- 948
DB 1242 NKLSELPPHELLTPVLGQSVPLGMGYSSDNGKITQEVIVSGADGPDNPSIGLIYTYLED 1301
QY 949 --NLDAAEHVNTLSNAFFQSLIEYNT-----TKESLSNLVAMKVQVYLAQFS-----T 996
DB 1302 LYNHVRMEGTLNSQ-NLRQLLENSVSSCFLTEQSIQKLLSEAEKRPYQSITLHQHULT 1360
QY 997 GLNTITDASKVVLVSTALDET 1018
DB 1361 GLPTIADAT--LSLLSVGLPGT 1380
```

RESULT 14

Q9RPH1

ID Q9RPH1

PRELIMINARY; PRT; 3223 AA.

```
AC Q9RPH1;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAY-2000 (TREMREL. 13, Last annotation update)
DE EHEC factor for adherence.
GN EFAL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E45035;
RX MEDLINE=20111988; PubMed=10652089;
RA Nicholls L., Grant T.H., Robins-Browne R.M.;
RT "Identification of a novel genetic locus that is required for in vitro
RT adhesion of a clinical isolate of enterohaemorrhagic Escherichia coli
RT to epithelial cells.";
RL Mol. Microbiol. 35:275-288(2000).
DR EMBL; AF159462; AAD49229.2; -.
DR InterPro; IPR001917; NHCtransf_2.
DR InterPro; IPR000169; SHprot_acsite.
DR TIGRFAMs; TIGR01586; yopT_cvs_prot; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS_1.
SQ SEQUENCE 3223 AA; 365949 MW; CAFD59F69242D07A CRC64;

Query Match 7.4%; Score 384; DB 2; Length 3223;
Best Local Similarity 19.8%; Pred. No. 5.3e-07;
Matches 254; Conservative 203; Mismatches 433; Indels 392; Gaps 55;

QY 2 NLVNAQLOKQVYKFRIOEDYVAILNALSEYHNMSSE-----VVEKYLK-LKDINN 54
DB 226 NFVQKISLSDELTKY---ANEIIEIKRIMGEYNLLPDKNSRNGKLKQKQADLLKIIME 282
QY 55 LTDNLYNTYKSGRNKALKKFKELYTMEVLEKN---NSLTPVEKNLHFIMIGQINDTA 111
DB 283 DTSVTENTFKN-----IEMAITDKREYYSHTVDIEKNIHAIWAGSPESI 329
QY 112 INYINQWQDVNSDYTVKFDYDSNAFLINTLTKKTIVESATNTNTLESFRENLDNPEFDYK 171
DB 330 SDYIKTLTKYKFTYLLWVDEKAFGAAKFTSVLKQIAFDLACRTIQNTPOKNDIFNL 389
QY 172 Y---RKR-----NEIYDQKHIDY-- 189
DB 390 YNEIRKKYNNPSSGOQYLNKRLRELYATYQIKSTPLKHMFSFFLENMILKQDNFFNYCI 449
QY 190 YKSQIEENPEFTID--NIKTYLSNE---YSKDEALNKYIBESL-----NKIT 233
DB 450 VKGVTIENDELRLNLYKNVVK--LSDDDIGNYQKTIENDKRVKLLIIDLQKQFGENRIS 507
QY 234 ANNGDIRLEKFADEDLVRLYNQELVERWNLAAASDILRLSMKEDGGVYLDVILPCI 293
DB 508 IKDVNSLTSLSKSENN---HNYQTMLLRWNPYPAASDILRMVILKEHGGIYTDMPAY 564
QY 294 QPDLFINKPSDITNTSWEMIKL-----EAMKYKEYIPGYTSKNFMDLD--EEVQSPFE 347
DB 565 SKQVIFKIMQMGNDNRFLDLKLRRAISDGLVRY-----VNNQNIDEVVNEYI----- 613
QY 348 SALSSKSKSEIFLPDLDIKVSPLEVKIAPANNSV-----INQ 385
DB 614 ----SDADKNIKKILTEISKMPEDSIFTKINTRIPROTMPIRLRYHLWPGWNRGLNG 669
QY 386 ALISLKDSYCSDLVI---NQIKNRYKILNDNLNPSINEGTDFNTMTKIFSDKLASINED 442
DB 670 FMLSHKSGEVIDAVIAGQNAVRELRIRDNHSEIY---FKQT-----DELSLDPD 720
QY 443 NMFMIKITNYLKVG-----FAPDVRSTINLSGP----- 471
DB 721 KIGGIL-VKKYLSGSLFSKFRQDTIPEALSTLQISGPDLIQKMLQFRSRGVLEBEFI 779
QY 472 ----GVY-----TGAYQ-----DILMFKN 488
```

Db 780 NERKLSKAYIGVYKTTGKGYDMLTPESIGVNDVTPADESTWICIGKRCVDDFLFKQVS 839  
Qy 489 TNIHLLPELRNFEPPKTKISQTEITSLASFQAR-AKQSFEEYKKGYPEGALGEDD 547  
Db 840 T-----LKTENLPFLTKIDTDTPFSQSTTKKDLQKKIQDLTVRYNE--LIDSS 889  
Qy 548 NLDFQANTVLDK-----DYVSKKILSMKTRNKEYIHYIYVLOQDKISYBASCNL 597  
Db 890 TIDFKNLYEIDQMLHMLMNDIAKRSLSQVQIAEKIRMT-----IPVDNIINI 943  
Qy 598 FSKDPYSSILYQKNIEGSTATYYVADAEIKEDYRIPYQISNKRNIKLTFIGHGKSE 657  
Db 944 YPD-----LHKNDNDLSMSIKGLAS--NPHTKINILYSNKTEHNI----- 983  
Qy 658 FNTDFTANLDVDSLSSEETILNLAADISPKYIEINLLGNCMFSYIYAEETYPCKLL- 716  
Db 984 FIKDLFS---FAVMENELRDIINNMSKDKTP-----ENWEGRVML 1020  
Qy 717 -----LKIKDRVSELMPSISQDSITVSANOYE---VRINEEG-KREILDHS--GKWINKE 765  
Db 1021 QRYLELKMCKDHLSQLSQSQEANEFLISTFIYENDFLREKIEAVKKNMSHELYFEKIKE 1080  
Qy 766 ESIIDKISKEY-----ISFNPKENKIIVSKYLHELSTLLQEIIRNNANSSDIDL 815  
Db 1081 QNTWODLSLTKQKQLIKALKESIGNTEKD---SHYDRLLDAPFK--KHNE-----I 1128  
Qy 816 EKKWMLTECEINVAS-----NIDROIVEG---RTEAKNLTSD-----SINY 854  
Db 1129 HNKIQRIDKDFKEYSRVAIHNIDKVIKGTDLRHYEGYVFSIDNTLSRYTLHLGIGTG 1188  
Qy 855 IKNEFKLIBSIDSYLKHONGLDSDHSPISDETSKTENGFRINPINKETGNSPIETE 914  
Db 1189 VTEENLLPAPASSLINILKE-----HYNEDEISAKLPLAYDY-ILNKKESSIPVEIL 1241  
Qy 915 KEIFS-----EYATHISKEISNIKOTIFDNVNGKLVKKV----- 948  
Db 1242 NKLSELPPHELLTPVLQSVNPLGMGYSSDNGKITEQVIVSGADGPDNPSIGLIVTYLED 1301  
Qy 949 --NLDAHEVNTLSNAFFQSLIEYNT-----TKESLNSLVAMKVQVYQAIFS-----T 996  
Db 1302 LYNIHVRMEGTLNSQ-NLRQLLENSVSSCFLTEQSINKLSAEAKRPYQSUTEIHOHLT 1360  
Qy 997 GLNTITDASKVVELVSTALDET 1018  
Db 1361 GLPTTADAT--LSLLSVGLPGT 1380

RESULT 15  
Q9RM48  
ID Q9RM48 PRELIMINARY; PRT: 3223 AA.  
AC Q9RM48;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Lymphostatin.  
GN LiFA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E2348/69;  
RX MEDLINE=20187489; PubMed=10722613;  
RA Klapproth J.M.A., Scaletsky I.C.A., McNamara B.P., Lai L.C.,  
RA Maltrom C., James S.P., Donnenberg M.S.;  
RT "A large toxin from pathogenic Escherichia coli strains that inhibits  
RT lymphocyte activation."  
RL Infect. Immun. 68:2148-2155 (2000).  
DR EMBL; AJ133705; CAB55629.1; -  
DR InterPro; IPR001917; NHRtransf.2.  
DR InterPro; IPR000169; SHprot\_acsite.

DR InterPro; IPR006473; Yop T.  
DR TIGRFAMs; TIGR01586; YOP1\_Cvs\_prot; 1.  
DR PROSITE; PS00599; AA TRANSFER\_CLASS\_2; 1.  
DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 1.  
SQ SEQUENCE 3223 AA; 365963 MW; 03B45CEDDE7938C1 CRC64;  
Query Match 7.3%; Score 381; DB 2; Length 3223;  
Best Local Similarity 19.7%; Pred. No. 7e-07;  
Matches 253; Conservative 204; Mismatches 433; Indels 392; Gaps 55;  
Qy 2 NLVNAQLOKQVYKFRIOEDYVAILNALSEYHNMSESS-----VVEKYLK-LKDINN 54  
Db 226 NFVQKISLSDLKTKY---ANEIIEIKRIMGEYNLLPDKNRNGUKLQKQADLLKIME 282  
Qy 55 LTDNYLNTYKSGRNKALKKFKYLTMBVLKLN---NSLTPVENKLNLFITWGGQINDTA 111  
Db 283 DTSVTENTFKN-----JEIAITDKREYYSHTVDIEKNIHAIWAGPPPSI 329  
Qy 112 INYINQWKDVNSDYTVKPYVDSNAFLINTLKKTIIVESATNNNTLESFRENLDNPEFYKPF 171  
Db 330 SDYIKTFLTKYKFTYYLWVDEKAPGAAKFTSVLKQIAFDLACRTIQQNTPKQKIDFINL 389  
Qy 172 Y---RKR-----MEIYDKQKHIDY--- 189  
Db 390 YNEIRKCKYNNPSSGOQYLNKRLRYATYQKISTPLKHMFSFFLENMIKLDQDNFFYCI 449  
Qy 190 YKSOIETENPEFTID--NIKITYLSNE---YSKDEALNKYIEBSL-----NKIT 233  
Db 450 VKGVTIEINDELRIYLNKVKIK--LSDDDIGNYQKTIINDKRVKGLIIDLQKQFENRIS 507  
Qy 234 ANNGNDIRNLEKPADDELRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILRQI 293  
Db 508 IKDVNSLTSLSKSENN---HNYQTEMLRWVYPAASDLLRWYLKHEGGIYTTDMPAY 564  
Qy 294 QPDLFKSINKPDSITNTSWEMIKL-----EAIMKTYEIPGYTSQNFMDL-EVQORSPE 347  
Db 565 SKQVFIKMMQTNQGNRFLDLKLRRAISDGLRY-----VNNQNIDEVNRYNEI--- 613  
Qy 348 SALSSEKSKSEIFLPLDDIKVSPLEVKIAPANNSV-----INQ 385  
Db 614 ---SDAKNIKKILTIEISKMPEDSIPTKINTRIPROTMPILRRYHLWPDGWNIRGLNG 669  
Qy 386 ALISLSDSYCSDLVT---NQIKNRYKILNDNLNPSINEGTDFNTMTKIFSCKLASINBD 442  
Db 670 FMLSCHKSEVIDAVIAGQNAVRELRIRDNIHSEIY---PKQT-----DELSLSPD 720  
Qy 443 NMFMKIKTNILKVG-----PAPDVRSTINLSGP----- 471  
Db 721 KIGGIL-VKKYLSGSLFSKFRQDTIIEALSTLQISGPDLIQKQMLQFFRSRGVLGEEFI 779  
Qy 472 -----GVY---TGAYQ-----DLLMFKONS 488  
Db 780 NERKLSKAYIGVYKTTGKGYDMLTPESIGVNDVTPADESTWICIGKRCVDDFLFKQVS 839  
Qy 489 TNIHLLPELRNFEPPKTKISQTEITSLASFQAR-AKQSFEEYKKGYPEGALGEDD 547  
Db 840 T-----LKTENLPFLTKIDTDTPFSQSTTKKDLQKKIQDLTVRYNE--LIDSS 889  
Qy 548 NLDFQANTVLDK-----DYVSKKILSMKTRNKEYIHYIYVLOQDKISYBASCNL 597  
Db 890 TIDFKNLYEIDQMLHMLMNDIAKRSLSQVQIAEKIRMT-----IPVDNIINI 943  
Qy 598 FSKDPYSSILYQKNIEGSTATYYVADAEIKEDYRIPYQISNKRNIKLTFIGHGKSE 657  
Db 944 YPD-----LHKNDNDLSMSIKGLAS--NPHTKINILYSNKTEHNI----- 983  
Qy 658 FNTDFTANLDVDSLSSEETILNLAADISPKYIEINLLGNCMFSYIYAEETYPCKLL- 716  
Db 984 FIKDLFS---FAVMENELRDIINNMSKDKTP-----ENWEGRVML 1020  
Qy 717 -----LKIKDRVSELMPSISQDSITVSANOYE---VRINEEG-KREILDHS--GKWINKE 765  
Db 1021 QRYLELKMCKDHLSQLSQSQEANEFLISTFIYENDFLREKIEAVKKNMSHELYFEKIKE 1080



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 19:25:02 ; Search time 51.25 Seconds  
(without alignments)  
5265.077 Million cell updates/sec

Title: US-09-126-816B-6\_COPY\_1\_1700  
Perfect score: 8677  
Sequence: 1 MNLVKNKAQLQKVVYKPRIQ.....YLYGIDRYNVKVIAPNLYT 1700

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6730.5	77.6	2366	17 AAR95011	C. difficile toxin
2	6730.5	77.6	2366	19 AAW68388	Clostridium diffi
3	4458	51.4	2710	17 AAR95016	C. difficile toxin
4	4458	51.4	2710	19 AAW68387	Clostridium diffi
5	2780	32.0	546	20 AAY33700	C. sordellii letha
6	3398	4.6	1979	21 AAB18171	Plasmodium falcipa
7	357.5	4.1	5024	22 AAG82935	S. epidermidis ope
8	348	4.0	10182	23 ABP38314	Staphylococcus epi
9	337	3.9	2485	21 AAB18172	Plasmodium falcipa

10	314.5	3.6	2496	21 AAB18222	Plasmodium falcipa
11	313	3.6	3973	22 AAB18253	Plasmodium falcipa
12	310	3.6	2184	22 AAE00425	P. falciparum telo
13	309	3.6	1639	19 AAW54145	P. falciparum synt
14	309	3.6	1639	23 AAE29345	Plasmodium falcipa
15	306.5	3.5	1516	21 AAB18195	Plasmodium falcipa
16	303.5	3.5	1802	21 AAB18217	Plasmodium falcipa
17	302	3.5	1254	11 AAR07503	Merzoite apical-en
18	302	3.5	1254	18 AAW24575	Merzoite apical-e
19	297.5	3.4	1308	21 AAB18167	Plasmodium falcipa
20	297	3.4	1654	6 AAP50777	Sequence of the P1
21	296	3.4	2010	21 AAB18218	Plasmodium falcipa
22	293	3.4	2539	21 AAB18198	Plasmodium falcipa
23	288.5	3.3	1948	23 ABP73774	Candida albicans e
24	287.5	3.3	1844	21 AAB18250	Plasmodium falcipa
25	287	3.3	2380	21 AAB18315	Plasmodium falcipa
26	287	3.3	3433	18 AAW22017	Uroophin. Homo sa
27	284	3.3	1881	23 ABP73809	Candida albicans e
28	283	3.3	2295	21 AAB18180	Plasmodium falcipa
29	282	3.2	2681	24 ABJ19025	Pathogen specific
30	281	3.2	1436	21 AAB18199	Plasmodium falcipa
31	281	3.2	3696	23 ABP40235	Staphylococcus epi
32	280	3.2	1817	21 AAB18301	Plasmodium falcipa
33	275	3.2	1712	21 AAB18205	Plasmodium falcipa
34	274	3.2	4134	20 AAY31946	Plasmodium falcipa
35	272	3.1	980	21 AAB18294	Plasmodium falcipa
36	271	3.1	2633	22 ABG06505	Novel human diagno
37	271	3.1	7201	22 ABB71136	Drosophila melanog
38	268.5	3.1	1847	21 AAY52002	M. jannaschii MJ14
39	268.5	3.1	1847	21 AAY51631	M. jannaschii MJ14
40	268	3.1	1477	20 AAY19981	B. burgdorferi ant
41	268	3.1	1494	20 AAY19980	B. burgdorferi ant
42	268	3.1	2663	22 AAM39097	Human polypeptide
43	267.5	3.1	2573	21 AAB18234	Plasmodium falcipa
44	267	3.1	1817	21 AAB18255	Plasmodium falcipa
45	266.5	3.1	1121	21 AAB18241	Plasmodium falcipa

ALIGNMENTS

RESULT 1  
AAR95011  
ID AAR95011 standard; Protein; 2366 AA.  
XX  
AC AAR95011;  
XX  
DT 08-JUL-1996 (first entry)  
XX  
DE C. difficile toxin B.  
XX  
KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;  
KW diarrhoea; therapy; diagnosis; vaccine.  
XX  
OS Clostridium difficile VPI strain 10463 (ATCC 10463).  
XX  
PN WO9612802-A1.  
XX  
PD 02-MAY-1996.  
XX  
PF 23-OCT-1995; 95WO-US13737.  
XX  
PR 07-JUN-1995; 95US-0480604.  
PR 24-OCT-1994; 94US-0329154.  
PR 16-MAR-1995; 95US-0405496.  
PR 14-APR-1995; 95US-0422711.  
XX  
(OPHI-) OPHIDIAN PHARM INC.  
PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
PI Williams JA;  
XX  
DR WPI; 1996-230603/23.

DR N-PSDB; AAT29247.  
 XX Fusion proteins comprising non-toxin protein and part of toxin -  
 PT useful to form anti-toxins against Clostridium botulinum type A, and  
 PT C. difficile type toxins, and to treat C. difficile intoxication,  
 PT partic. diarrhoea  
 XX  
 PS Claim 36; Page 313-323; 434pp; English.  
 XX  
 CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the  
 CC product of the toxin B gene (AAT29247), is a cytotoxin associated  
 CC with diarrhoeic disease. It can be obtd. by expression in  
 CC transformed E. coli hosts of portions of DNA that together cover the  
 CC entire toxin B gene. Toxin B, and portions of it (see also AAR95012-13,  
 CC AAR95371-72 and AAR95018), pref. expressed as fusions to polyhistidine  
 CC affinity tags or maltose binding protein, are used to raise avian  
 CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.  
 XX  
 SQ Sequence 2366 AA;  
 Query Match 77.6%; Score 6730.5; DB 17; Length 2366;  
 Best Local Similarity 76.4%; Pred. No. 0;  
 Matches 1298; Conservative 191; Mismatches 210; Indels 1; Gaps 1;  
 1 MSLVNKAQQLQXMYVVFRIQDEYVAILNALAEYHNMSSESVVEKYKLKDINNLTDNYL 60  
 1 MSLVNRKQLEKXANVFRQDEYVAILDALAEYHNMSSESVVEKYKLKDINSLTDYI 60  
 61 NTYKSGRNKALKKFKXEYITMEVLELKNLSLTPVEKNLHFWIGGQINDTAINYNQWKD 120  
 61 DTYKSGRNKALKKFKXEYITMEVLELKNLSLTPVEKNLHFWIGGQINDTAINYNQWKD 120  
 121 VNSDYTVKVFYDSNAFLINTLKTIVTESATNTLESFRENLPDPEYKFKRMEIY 180  
 121 VNSDYNNVVFYDSNAFLINTLKTIVTESATNTLESFRENLPDPEYKFKRMEIY 180  
 181 DRQKHFDYKYSQIEENPEFIDNIIKTYLSNEYSKDLALNKYIEESLNKITTANNNDI 240  
 181 DRQKNFINYKAQREENPELLIDDIKTYLSNEYSKEIDELNTYIEESLNKITTQSGNDV 240  
 241 RNLEKPADEBLVRLNQEILVERWNLAASDILRISMLKEDGGYVLDVDPGIDPLFKS 300  
 241 RNPEFKNGESFNLYEQELVERWNLAASDILRISALKEIGWYLDVDMPLGIDPLFS 300  
 301 INKPSDITWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQSPESALSKSKSEIF 360  
 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQSPESALSKSKSEIF 360  
 361 LPDDDIKVPLEVKIAFANNVINOALISLKDSCYSDLVINOIKNRYKILNDNLNPSINE 420  
 361 SSGDMEASPLEVKIAFNSKGIINOGLISVKDSYCSNLIVKQIENRYKILNLSNLPASE 420  
 421 GTDFNTMTKIFSDKLASISNEDNMPMIKITNYLKYGFAPDVRSTINLSGPGYTTGAYD 480  
 421 DNDFNNTTTFIDSIMAEANADNGRPMBELGKYLVRGFPDPVKTTLNLSGPEAYAAQD 480  
 481 LLMFKDNSTNIHLLEPELRFNFPPTKISQLTQEBITSLWSFNQAPAKSQPEYKKGYPE 540  
 481 LLMFKGSGNIIHLLEADLRFNFEISKTNISQSTEQEASLWSFDDAKAQPEYKRNYPE 540  
 541 GALGEDNDLDFACNTVLDKDYVSKITLSMKTRNKEYIHVIYVLOQDKISYEASCNLFSK 600  
 541 GSGLEDNDLDFSNIVVDKEYLLEKISSLARSEYGIHVIYVLOQDKISYEASCNLPAK 600  
 601 DPVSSILYQKNEGSTAYVYVADAEIKEDIRIPYQISNKNRKLTFIIGHKSEFNT 660  
 601 TPYDSVLFKNIEDSEIAYVNPGEIGQIDKYKIPSIISDRPKIKLTFIGHKDEFT 660  
 661 DTFANLVDLSLSEIETILNLAADISPKYIEINLILGCMNMFYSIYAETYPCKLLIKK 720  
 661 DIPAGFDVLSLSEIAAIDLAKEDISPKSIEINLILGCMNMFYSINVEETYPCKLLIKK 720  
 721 DRVSELMPSISQDSITVSANQVEVRINEBGRKILDHSGKWINKEBSIIKDISSEYISF 780

RESULT 2  
 AAW68388  
 ID AAW68388 standard; Protein; 2366 AA.  
 XX  
 AC AAW68388;

721 DKISELMPSISQDSITVSANQVEVRINEBGRKILDHSGKWINKEBSIIKDISSEYISF 780  
 781 NPKENKIIVKSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDRIVEGR 840  
 781 NPKENKIIVKSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDRIVEGR 840  
 841 IEEAKNLTSDSINYIKNEFKLIESISLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900  
 841 IEEAKNLTSDSINYIKNEFKLIESISLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900  
 901 INKETGSGIFTEKEIFESEYATHISEISNIKOTIFDNVNGKLVKKNLDAHVNTLIN 960  
 901 INKETGSGIFTEKEIFESEYATHISEISNIKOTIFDNVNGKLVKKNLDAHVNTLIN 960  
 961 SAPPFQSLIEYNTTKESISNLSVAMKVQYVQALFSTGLNTITDASKVELVSTALDETID 1020  
 961 SAPPFQSLIEYNTTKESISNLSVAMKVQYVQALFSTGLNTITDASKVELVSTALDETID 1020  
 1021 LPLTSEGLPIATIIDGVSLGAATKELSETNDPLRQIEIAKIGIMAVNLTAATAVT 1080  
 1021 LPLTSEGLPIATIIDGVSLGAATKELSETNDPLRQIEIAKIGIMAVNLTAATAVT 1080  
 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGAFITLD 1140  
 1081 SSGIASGFSILLVPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGAFITLD 1140  
 1141 DKIMPDQDLVLEISDFNNNSITLQKCEIWRAGSGGHTLTDIDHFFSSPSITVRKPKWL 1200  
 1141 DKIMPDQDLVLEISDFNNNSITLQKCEIWRAGSGGHTLTDIDHFFSSPSITVRKPKWL 1200  
 1201 STYDVLNITKKEIDKPSKDLMLPNAFVFGVEMGTGPFSLDNDGKTLDRIDHVEG 1260  
 1201 STYDVLNITKKEIDKPSKDLMLPNAFVFGVEMGTGPFSLDNDGKTLDRIDHVEG 1260  
 1261 QFYRYFAFIADALITLKPREDTNVRLNDGNTRSPFVITTEQIRKNLSYFSGG 1320  
 1261 QFYRYFAFIADALITLKPREDTNVRLNDGNTRSPFVITTEQIRKNLSYFSGG 1320  
 1321 GYSLSLSPYNNIDLNLVENDTVIDVNVVKNITTESDEIQKGLIENILSKUNIEDN 1380  
 1321 GYSLSLSPYNNIDLNLVENDTVIDVNVVKNITTESDEIQKGLIENILSKUNIEDN 1380  
 1381 KIILNHTINFVGDINESRIFSLTFSLEEDINIIIEIDLVSYSKILLSGNCKLIENS 1440  
 1381 KIILNHTINFVGDINESRIFSLTFSLEEDINIIIEIDLVSYSKILLSGNCKLIENS 1440  
 1441 SDIOQKIDHIFNGEHOQYIFVSYIDNETKNGFIDYSKGLFTAEPNSESIIIRNIYMP 1500  
 1441 SDIOQKIDHIFNGEHOQYIFVSYIDNETKNGFIDYSKGLFTAEPNSESIIIRNIYMP 1500  
 1501 DSNLFIYSSKDLKDIRIINKGDKVLLIGNYFKDKMKSLSFTIEDTNTIKUNGVLDBN 1560  
 1501 DSNLFIYSSKDLKDIRIINKGDKVLLIGNYFKDKMKSLSFTIEDTNTIKUNGVLDBN 1560  
 1561 GVAQITLKEMNAKSALNTSNIAMPELESINIKIEFNNLDNIEFILDTNFIISGNSIG 1620  
 1561 GVAQITLKEMNAKSALNTSNIAMPELESINIKIEFNNLDNIEFILDTNFIISGNSIG 1620  
 1621 QPEFLICDCKDKNIQPYFINFKIKETSYTLVGNRQNLIVEPSYHLLDSDSGNISSTVNFQK 1680  
 1621 QPEFLICDCKDKNIQPYFINFKIKETSYTLVGNRQNLIVEPSYHLLDSDSGNISSTVNFQK 1680  
 1681 YLYGIDRVNKKVIAPNLYT 1700  
 1681 YLYGIDRVNKKVIAPNLYT 1700  
 1699 YLYGIDRVNKKVIAPNLYT 1699  
 1699 YLYGIDRVNKKVIAPNLYT 1699



XX	07-DEC-1998	(first entry)
DT	Clostridium difficile toxin B.	
XX	Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;	
XX	pseudomembranous enterocolitis.	
KW	Clostridium difficile.	
XX	WO9808540-A1.	
XX	05-MAR-1998.	
PD	28-AUG-1997;	97MO-US15394.
PF	28-AUG-1996;	96US-0704159.
XX	(OPHI-) OPHIDIAN PHARM INC.	
PA	Thalley BS, Williams JA;	
XX	WPI; 1998-230234/20.	
DR	N-PSDB; AAV30561.	
XX	Host cell containing recombinant expression vector encoding	
PT	Clostridium botulinum type B or E toxin - useful to treat humans	
PT	and other animals at risk of intoxication with clostridial toxin	
XX	Example 18; Page 241-249; 428pp; English.	
PS	This is the amino acid sequence of Clostridium difficile toxin B,	
XX	deduced from the coding region (see AAV30561) of the toxin B gene.	
CC	Fragments of the toxin B gene have been cloned into various	
CC	prokaryotic expression systems, and assessed for the ability to	
CC	express recombinant toxin B protein in E. coli. It would be	
CC	advantageous to use simple and inexpensive prokaryotic expression	
CC	systems to produce and purify high levels of recombinant toxin B	
CC	for immunisation purposes. The invention specifically relates to	
CC	recombinant proteins derived from Clostridium botulinum toxins	
CC	(see AAW68389-400) and their use as immunogens for the production of	
CC	vaccines and antitoxins.	
XX	Sequence	2366 AA;
SQ	Query Match	77.6%; Score 6730.5; DB 19; Length 2366;
	Best Local Similarity	76.4%; Pred. No. 0;
	Matches 1298; Conservative 191; Mismatches 210; Indels 1; Gaps 1	
Qy	1	MNLVYKQLQKVYVKFRIQDEYVAIINALVEEYHNMSSESVKEYLKDINLLTDNYL 60
Dd	1	MSLVNRKQLEKANVRFRFTQDEYVAIIDLAEYHNMSSENTVVEKYLKLDINSLTDIYI 60
Qy	61	NTYKKSGRNKALKKPKLYLTMEVLKKNLSLTPVEKNLHFVIWGQINDTAINYNQWKD 120
Dd	61	DTYKKSGRNKALKKFEYLTVTEVLKKNNSLTPVEKNLHFVIWGQINDTAINYNQWKD 120
Qy	121	VNSDYTKVFVDNSNAFLINTLUKTIVESATNWTLESFRENLMNDPFDYNNKFKRMEIYI 180
Dd	121	VNSDYNVNVPVDSNAFLINTLUKTIVESAINDTLFSFRENLMNDPRFDYNNKFFRKMQIIY 180
Qy	181	DKQHFDYYKSQTREENPEFIIDNIITKYLISNEYSKDLKALKNYIEESLINKITANNNDI 240
Dd	181	DKQNFINFYKAQRENPELIIDDIVKTYLSNEYSKEDIELMTYIEESLINKITQNGNDV 240
Qy	241	RNLSEKFADEDILRVLYNQBELVERWNLAASADILIRISMLKEDGGVYLDVILPGIQDFLFS 300
Dd	241	RNFEPKNGESFNLYEQELVERWNLAASADILIRISALKEIGMYLDDVMLPGIQDFLPES 300
Qy	301	INKDPSTINTSWEMIKLEAIWMKYKEYIPGYTSKNFMDLDEEVORFEGALSCKSKSEIF 360
Dd	301	IEKPSSVTVDPWENTKLEAIWMKYKEYIPEYTSEHFDMLDEEVQSSEFVSGLASKSKSEIF 360

```

Db      1441 NHIIQQIDYGFNSLQKNIPIYSFVDSKENGFGFINGSTKEGLFVSELPDVLISKVYMD 1500
QY      1501 DSNLFIYSKDKJOTRIKNGDKVLLIGNYFKDDMKVSLSFPTIEDTNTIKUNGVLVDN 1560
Db      1501 DSKPSFGYSSNNLKOVKITKONVNLTGYLKKDDIKISLSLTQDEKTIKLSNVHLDS 1560
QY      1561 GVAQILKFNNAKSALNTSNLSMNFLESINIKNIFYNNLDPNLEFILDNFIISGNSIG 1620
Db      1561 GVAEILKFNMR-KGNTNTSDLSMFLSNMKSIFVNFQSNIKFILDANFIISGTTISG 1619
QY      1621 QFELICDDKNIOPIYFINKIKETSITLVGNRQNLIVEPSYHLDDSGNISSTVINFSOK 1680
Db      1620 QFELICDENDIOPIYKNTLETNTLVGNRQNWIVEPNVDLDDSGDISSTVINFSOK 1679
QY      1681 YLYGIDRYNKVITIANLYT 1700
Db      1680 YLYGIDSCVKNVISPNIYT 1699

```

RESULT 3

```

ID AAR95016 standard; Protein; 2710 AA.
XX
AC AAR95016;
XX
DT 08-JUL-1996 (first entry)
XX
DE C. difficile toxin A.
XX
KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;
KW vaccine; diarrhoea; diagnosis; therapy.
XX
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX
PN W09612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US13737.
XX
PR 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX
DR WPI; 1996-230603/23.
DR N-PSDB; AAT29248.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin -
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX
PS Claim 63; Page 290-302; 434pp; English.
XX
CC Clostridium difficile VPI strain 10463 toxin A (AAR95016), the
CC product of the toxin A gene (AAT29248), is a potent cytotoxin that
CC plays a direct role in damaging gastrointestinal tissues and is
CC associated with diarrhoeic disease. It can be obtd. by expression in
CC transformed E. coli hosts of portions of DNA that together cover the
CC entire toxin A gene. Toxin A, and portions of it (see also
CC AAR95014-15 and AAR95017), pref. expressed as fusions to polyhistidine
CC affinity tags or maltose binding protein, are used to raise avian
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodrn.
XX
SQ Sequence 2710 AA;

```

```

Query Match      51.4%; Score 4458; DB 17; Length 2710;
Best Local Similarity 50.7%; Pred. No. 1.6e-226;
Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

QY      1 MNLVNAQLQRMVYKFRIOBDEYVAILNAALEEYHNMESSVVEKYLKDKOINLNTDYL 60
Db      1 MSLISKEELIKLAY-SIRPRENEYKITLTNDLEYNKLTNNNENKYLQKKLNEIDVPM 59
QY      61 NTYKSGRGNKALKPKKEYLTMEVLELKNSTLTPVEKNLHFHWIGQINDTAINYNQWKD 120
Db      60 NKYTSSNRNALSNLKDKILKEVILIKNSNTPVEKNLHFHWIGGEVSDIALEYKQWAD 119
QY      121 VNSDYTVFVVDNAFLNTLTKTIVESATNTNTLESFRENLDNDFEDYKFKYKWEIITY 180
Db      120 INAEYNIKLVWDEAFVNTLKALVSSSTTEALQLLEEBEQNPQDNMKFYKKWEFIY 179
QY      181 DKQHFIDYKYSQIBENPEFIIDNIITKYLSEYKSKOLEALNKYTEESLNKITTANNNDI 240
Db      180 DRQKEFINYKSKINQKPTVPTIDDIKSHLVSEYNRDETVELSYRTNLSRKINSNHGDI 239
QY      241 RNLEKFADEDLVRLYNQBELVERWNAAASDILRISMLKEDGGVYLDVILPGIQDPLPKS 300
Db      240 RANSLFTEQELLNIYSQELLNRGNLAAASDIVRLALKNFVGGVYLDVMDLPGIHSDFKT 299
QY      301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQRFESALSSKSKSEIF 360
Db      300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKDDQDKQNFKLIIESKSEIF 359
QY      361 LPDDIKVPLEVKIAFANNVINOALISLKDSCYSDLVINQIKNRYKILNDNLNPSINE 420
Db      360 SKLENLVSDLEIKIAFALGVSINOALISKQSYLTNLVIEQVKRYQFLNQHLPATES 419
QY      421 GTDFNTTKIFSDKLASISNEDNMFMKIKNYLVKGFADPVRSTINISGPGVYTGAYOD 480
Db      420 DNNFTDTTKIFHDSLFNSATAENSMLTKIAPYLVQVGFMPPEARSTISLSPGAYASAYD 479
QY      481 LLMFKDNSTNIHLLPELNFEEPKTKISQTEQITSLMSFNOARAKOSEEYKGYFE 540
Db      480 FINQENTIEKTLKASDLIEPKFPENNLSQTEQEINLSWSPDOASAKYQEKYVRDVTG 539
QY      541 GALGEDNDLPAQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIYVLOQDKDISYEASCNL 597
Db      540 GSLSEDNGVDNFKNATLDKNVLLNNKIPSNVVEEAGSKNYVHYIILQOGDDISYEATCNL 599
QY      598 FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPYQISKNKNTKLTFIGHGK 655
Db      600 FSKNPKNSIIQRNM--NESAKSYFLSDDGESILLELNKYRIPERLKNKEKVKVTFIGHGK 657
QY      656 SEFNFTDTFANLDVDSLSEIETILNLAKADISPKYIEINLGCNMFSYIYAETYPGKL 715
Db      658 DEFNTSEFARLSVDLSLSNEISFELDTIKDISPKNVEVNLGCNMFSYDFNVEETYPGKL 717
QY      716 LLKIKDRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWKINKEESIIDKSK 775
Db      718 LLSIMDKITSLPDVNKNSITIGANQYEVRIINSEGRKELLASHGKWKINKEEAIMSDLSK 777
QY      776 EYISNPENKNIYVSKYLHELSTLLOQIRNANSSDIDLEKQVMLTCEINVASNIDRQ 835
Db      778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSPDTKTLNKLNIIESIGDY 837
QY      836 IVEGRIEBAKNLTSDSINYIKNEPKLIESISDSLYDLKHQGLDDSHFISFEDISKTEG 895
Db      838 IYVEKLEVPKNIHNSIDDLIDFENLLENVSDLEYELKUNLNDKYLISFEDISKNST 897
QY      896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIMKDTIPDNVNGKLVKKVNLDAAE 955
Db      898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQLDHTSQ 957
QY      956 VNTLNSAFIOQLSEYNTTKESLNSLVAMQVYLAQLFSTGLNTITDASKVVELVSTAL 1015
Db      958 VNTLNAAPFIOQLSIDYSSNKVDLNDLSTSVQVLAQLFSTGLNTIYDSIQVLNLSNAV 1017
QY      1016 DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQRETEAKIGIMAVNLTAAS 1075

```

Db 1018 NDTINVLPITTEGIPVTSTLDGILNGLAAIKELLDEHDPLLKKEAKVGLAINMSLSI 1077  
 Qy 1076 TAITVSALGIASGFSLLVPLAGISAGIPSLVNNELILODKATKVIDYKPHISLAETEGA 1135  
 Db 1078 AATVASIGIGAEVITFLPFIAGISAGIPSLVNNELILHDKATSVVYFNHLSSEKKYGP 1137  
 Qy 1136 FTLLDDKIIMPDDLVLSIDFNNSNITLKGCEIWRAGGSGHTLTDDIDHFFSSPSITY 1195  
 Db 1138 LKTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGSGHTVTGNIDHFFSSPSISS 1197  
 Qy 1196 RKPWLSIYDLNIKEKIDFSDKDLMLVLPNAPRVFGYEMGWTPGFRSLDNDGTGLDLRIR 1255  
 Db 1198 HIPSLSIYGAIGETENLDFSKIMWLPNAPSFRVFWETGAVPGLRSLDNDGTGLDLRIS 1257  
 Qy 1256 DHVEGOFVYRFAFTADALITKLKRYEDTVNRINLNDGTRSFIVPVITTEOIRKNLSYS 1315  
 Db 1258 DLYPGKFYWRFAFP--DYAITTLKPYEDTNIKIKLDKOTRPFIMPTITNEIRKNLSYS 1316  
 Qy 1316 FYGSGGSYLSLSPYNNMIDLNVENDTVIVDNDVNNKNTTIESDIOKGLIENILSKL 1375  
 Db 1317 FDGAGGTYSLLSYPISTINILSKDDLWIFNIDNEVREISIENTGKIKGLIKOVLSKI 1376  
 Qy 1376 NIEDNKIILNNHTINPYDINESNRPSISLTFSEIDINIIIDILVSKSYKILLSCNWK 1435  
 Db 1377 DINKKLIIGNQTIDPSGIDNDKDRYIFLTCBLDDKISLIIEINLVAKSYSLLSGDKNY 1436  
 Qy 1436 LIENSDDIOQKIDHIGFNGEHOKYIPYSYID-NETKYNCFIDYSKKEGLFTAEPFNSII 1494  
 Db 1437 LISLNSLTIEKINTLGLD---SKNTAAYNTDESNNKYFGAI-----SKTSQKSI 1483  
 Qy 1495 RNIIYWDPSNNL-----FIYSSKDL--KDIRINKGVDKLLIGNPKD---DMKVSLS 1541  
 Db 1484 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVPKDDINITGKYVDNNTDKSIDFS 1541  
 Qy 1542 FTIEDTNTIKANGVYLDENGVAQILKPMNNAKSALNTSNLMNFLESINIKNIFNNLDP 1601  
 Db 1542 ISLVSQNVKANGVLYNESVSYSLDFVKNSDGHHTNSFMNLFLDNISFWKLFGPE--- 1598  
 Qy 1602 NIEPILDTNPIISGNSISIQFELICDKDNIQIOPYFINFKIETSPLYVGNQNLIVERP 1661  
 Db 1599 NINFIYDKYFTLVGKNTLGVFEFCIDNNKNIDIFYGFEWKTSSKSTIFSNGRNVVVEPI 1658  
 Qy 1662 YHLDGSGNISSTVINFSQKLYCIDRYVANKVIAAPNLYT 1700  
 Db 1659 YN-PDTGEDISLDSFSEYPLGIDRYINKVLIAPDLYT 1696

## RESULT 4

AAW68387

ID AAW68387 standard; Protein; 2710 AA.

XX

AC AAW68387;

XX

DT 07-DEC-1998 (first entry)

XX

DE Clostridium difficile toxin A.

XX

KW Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;

XX

KW pseudomembranous enterocolitis.

XX

OS Clostridium difficile.

XX

PN WO9808540-A1.

XX

PD 05-MAR-1998.

XX

PP 28-AUG-1997; 97WO-US15394.

XX

PR 28-AUG-1996; 96US-0704159.

XX

PA (OPHI-) OPHIDIAN PHARM INC.

XX

PI Thalley BS, Williams JA;

XX

WPI: 1998-230234/20.

DR

N-PSDB; AAV30560.

XX

PT Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin

XX

PS Example 15; Page 220-230; 428pp; English.

XX

CC This is the amino acid sequence of Clostridium difficile toxin A,  
 CC deduced from the coding region (see AAV30560) of the toxin A gene.  
 CC Toxin A is a potent cytotoxin that plays a direct role in damaging  
 CC gastrointestinal tissues. Severe cases of C. difficile  
 CC intoxication result in pseudomembranous colitis. This would be  
 CC prevented by neutralising the effects of toxin A in the  
 CC gastrointestinal tract. Examples are provided of the production  
 CC of recombinant C. difficile toxin A in host cells and of the in  
 CC vivo neutralisation of toxin A by antibodies against recombinant  
 CC toxin A polypeptides. The invention specifically relates to  
 CC recombinant proteins derived from Clostridium botulinum toxins  
 CC (see AAW68389-400) and their use as immunogens for the production of  
 CC vaccines and antitoxins.

XX

SQ Sequence 2710 AA;

Query Match 51.4%; Score 4458; DB 19; Length 2710;

Best Local Similarity 50.7%; Pred. No. 1.6e-226;  
 Matches 871; Conservative 363; Mismatches 443; Indels 42; Gaps 15;

Qy 1 MNLVNAQLOKVVYKFRIOEDYVAIILNALAEYHNHMSSESVVEKYKLKDINNLTDNYL 60

Db

1 MSLISKEELIKLAY-SIRPRENEYKITLTNLDYNNKLTNNNNENKYLQKLKLNESIDVFM 59

Qy

61 NTYKSGRNKALKKPKYLTMEVLELKNLSLTPVEKNLHFIWIGGQINDTAINYQWKD 120

Db

60 NKYTSRRNALSNLKKOILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD 119

Qy

121 VNSDYTVKPVVDNAPLNTLTKTIVESATNTTLESFRENLDNDFYNNKFKYKMEIY 180

Db

120 INARYNIKILWYDEAPLVNTLTKAIVESSTTALQLEEEIQNPFDNNKFKYKMEIFY 179

Qy

181 DKQKHFDYKYSQIBENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITTANNNDI 240

Db

180 DRQKGFINYKSIQINKPTVTPTDIDIIKSHLVSEYNRDETVLSEYRNSLRKINSHGIDI 239

Qy

241 RNLEKFADEDLVRLYNQELVERWNLAASDILRIKMLKEDGGVYLDVILPGIQDPLFKS 300

Db

240 RANSLPTEQELNLYSQELLNRGNLAASDIVRLALKNFGVYLDVMDLPGIHSDLFKT 299

Qy

301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSKSKSEIF 360

Db

300 ISRPSISGLDRWEMIKLEAIMKYKYINNTYTSNFDFKLDQDKDNFKLIESKSEIF 359

Qy

361 LPLDDIKVSPLEVKIAFANNVINOALISLSDSYSDLVINOIKRYKILNDNLNPSINE 420

Db

360 SKLENLVSDLEIKIAFALGVSINOALISKQSYUTNLVIEQVKRYQFNLHNPALPIS 419

Qy

421 GTDFNTTWMKIFSDKLASINEDNNMFMKITTYLVKGFAPDVRSTINLSGGVYTGAYOD 480

Db

420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPPEARSTISLSPGAYASYD 479

Qy

481 LLMFKNSTNIHLEPELANFEPTKISQLTQEITSLSWSNQARAKSOFPEYKKGYPE 540

Db

480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTQEINSLSWSFDQASAKYQFEKYRDTYG 539

Qy

541 GALGDDNLDPAQNTVLDKDY-VSKKILSS--MKTRNKYIHYIYVQLQGDKLSYEASCNL 597

Db

540 GSLSEDNGVDFNKNLTALDKNYLLNKNIPSNVVEAGSKNYHYIYIQLQGDGDDISYEATCNL 599

Qy

598 FSKDPYSSILYQKNIEGSETAYYYVAD--ABIKEIDKYRIPYQISNKNKIKLTFIGHGK 655

600 FSKNPKNSIIQRM--NESAKSYFLSDGSEISLELNKYRIPERLKNKEKVVTFIGHGK 657  
 656 SERNTDTFANLDVDSLSSEIETILNLAADISPKYIEINLGNMPSYSIYAEETPGKL 715  
 658 DEFNTSEFARLSVDSLSNBSISGLFDIKDIDSPKNVEVNLGNMPSYDFNVEETPGKL 717  
 716 LKIKDRVSELMPISQDSITVSANQYEVINEEGKREITLDHSGKWINKKEESIIKDSSK 775  
 718 LLSIMDKITSTPDVNNKSIITIGANQYEVINEEGKREITLDHSGKWINKKEESIIKDSSK 777  
 776 EYISFNPKNKIIVKSKYLHEISLTLQRIINNANSSDIDLEKKVMLTECEINVAQNDRQ 835  
 778 EYIFDSIDNKLKAKSKNIPGLASISEDIKTLDDASVDPDKFILNKLNISSIGDY 837  
 836 IVEGRLEEAKNLTSDSINYIKNEPKLIESISDLYDLKHONGLDSDHPSFEDISKENG 895  
 838 IYKLEPKVKNIIHNSIDILDEFNLENVSDLEYELKKNLDEKYLISFEDISKENST 897  
 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAHE 955  
 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQDHTSQ 957  
 956 VNTLSAFTIQLSIEYNTTKESLSNLSVAMQVVAQLFSTGLNTITDASKVVELVSTAL 1015  
 958 VNTLSAFTIQLSIEYNTTKESLSNLSVAMQVVAQLFSTGLNTITDASKVVELVSTAL 1017  
 1016 DETIDLPLTSEGLPIATIIDGVSIGAAIKELSETNDPLLRQETEAIGIMAVNLTAAS 1075  
 1018 NOTINVLPTITGIPVSTILGGINLGAIAKELLDEHPLKKELEAKVGLVAINMSLSI 1077  
 1076 TAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKIDYFKHISLAETEGA 1135  
 1078 AATVASIVGIGAEVITFLPIAGISAGIPSLVNNELILODKATKIDYFKHISLAETEGA 1137  
 1136 FTLLDDKIIMPODVLSEIDFNNSITLGEKCEINRAEGSGHTLTDIDHFFSPSITY 1195  
 1138 LKTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGGSGHTVTGNIDHFFSPSITS 1197  
 1196 RKPWLSIYDLNLIKKEIDFSDKDLMLVLPNAPRVFGEYENGWTPGFRSLNDGDKLLDRIR 1255  
 1198 HIPSUSIYSAIGIETENLDFSKIMMLPNAPRVFGEYENGWTPGFRSLNDGDKLLDRIR 1257  
 1256 DHYEGOFYWRFAFIADALITKLKRYEDTVNRINLDGNTGRFIVPVIITEQIRKNLSYS 1315  
 1258 DLYPGKFRWFAFF-DYAITLTKPYVEDTNIKLKDKOTRFINPMTITNIRKNLSYS 1316  
 1316 FYGSGSYLSLSPYNNIDNLVENDTVIDVNVKNITIESDBIQGELIENILSKL 1375  
 1317 FDCAGGYSLLSSYPISNTINLSKDDLWIFNIDNEVREISIENTGTIKKGLIKDVLSKI 1376  
 1376 NIEDNKILNNHTINFYGDINESNREISLTFSTLEIDINIIIEIDLVSYSKILLSGNCWK 1435  
 1377 DINKNKLIIIGNQTDIFSGDIDNDRYIFUTCELDDKISLIIIEINLVAKSYSLLSGDKNY 1436  
 1436 LIENSSDIOQKHIDHGFNGEHQKIFYSVID-NETKNGFIDYSKKEGLFTAEFNSNII 1494  
 1437 LISLNSLTKINTGLD---SKNIAVNTDSNNKYPGAI-----SKTSQKSI 1483  
 1495 RNIYMPDSNNL-----FIYSSKDL--KDRIINKGDVKKLLIGNYFKD---DMKVSUS 1541  
 1484 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVMFKDDINTITGKYVYDNNNTDKSIDFS 1541  
 1542 FTIEDNTIKLNGVILDENGVAQILKFMNNAKALNTSLSNLFLESINIKIIFYNNLDP 1601  
 1542 ISLVSNQVKNGLYLNESVYSYLDVFNKSDGHNTSNFNLFLDNISFWLFGPE--- 1598  
 1602 NTEFIDLTNFIISGNSIGOFELICDKNIOFYFINFKIKETSITYLVGNRQNLIVEPS 1661  
 1599 NINFVIDKYFTLVGKTNLGVVEFICDNNKIDIFYGCEWTKSSKSTIFSGNGRNVVVEPI 1658  
 1662 YHLDSDGNTSSVTNFSQKLYGIDRYVKNVIAPIAPLYT 1700  
 1659 YN-PDTGEDISTSLDPSYPLGYIDRYINKVLIAPDLYT 1696

RESULT 5

AA33700

ID AA33700 standard; protein; 546 AA.

XX AC AA33700;

XX DT 18-JAN-2000 (first entry)

XX DE C. sordellii lethal toxin protein fragment.

XX KW Lethal toxin; immunotoxin; antitumor; glucosyl transferase; glycosylate;

XX KW GTPase; Ras; epidermal growth factor stimulated MAP-kinase; inhibition;

XX KW signalling pathway; cell-specific toxin; treatment; cancer.

XX OS Clostridium sordellii.

XX PN DE19802569-A1.

XX PD 09-SEP-1999.

XX PF 23-JAN-1998; 98DE-1002569.

XX PR 23-JAN-1998; 98DE-1002569.

XX PA (UYFR-) UNIV FREIBURG ALBERT-LUDWIGS.

XX PI Aktories K, Hofmann F;

XX DR WPI: 1999-509323/43.

XX N-PSDB; AA223800.

PT New fragment of the lethal toxin from Clostridium bacterium, useful for

XX PS treating cancer.

XX PS Claim 1; Page 7-9; 14pp; German.

CC This invention describes a novel fragment (I) of the lethal toxin (LT) of Clostridium sordellii which has antitumor activity. (I) is a glucosyl transferase that glycosylates, and thus inactivates, GTP(guanine triphosphate)ases, particularly Ras (an oncogenic product overexpressed in many tumors), resulting in inhibition of epidermal growth factor stimulated MAP-kinase signalling pathways. (I), particularly in the form of immunotoxins, are used as cell-specific toxins, particularly for treating cancer. When included in immunotoxins, (I) can be targeted to selected cells. Compared with the complete LT, (I) is smaller, so enters cells more easily, resulting in greater toxicity in the cytosol, is less likely to induce formation of (neutralizing) antibodies, and is more active than the holotoxin. This sequence represents the lethal toxin fragment described in the invention.

XX SQ Sequence 546 AA;

Query Match 32.0%; Score 2780; DB 20; Length 546;

Best Local Similarity 99.1%; Pred No. 8.2e-139;

Matches 541; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNLVNAQLQKQVYKFRIOEDYVAILNLEEEYHNMSESSVVEKYLKLDINNTDNYL 60

DB 1 MNLVNAQLQKQVYKFRIOEDYVAILNLEEEYHNMSESSVVEKYLKLDINNTDNYL 60

QY 61 NTKKSGRKNALKKPEKYLTMVEVLKNSLTPVEKNLHFVIGQINDTAINYNQWKD 120

DB 61 NTKKSGRKNALKKPEKYLTMVEVLKNSLTPVEKNLHFVIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVFVDSNAPLNTLKTIVESATNTNLTSPRENLDNPFYDKFKRKEIY 180

DB 121 VNSDYTVKVFVDSNAPLNTLKTIVESATNTNLTSPRENLDNPFYDKFKRKEIY 180

QY 181 DKQKHFDYKYSQIBENPEFIIDNIKTLYSNEYSKOLEALNKYTEESLNKITTANNNDI 240

DB 181 DKQKHFDYKYSQIBENPEFIIDNIKTLYSNEYSKOLEALNKYTEESLNKITTANNNDI 240



Db 674 SENILNDBE-----VCDLKRKLSLSESMKMKKEHD 706  
Qy 834 ROIVE-----GRIEAKNLTSDSINYIKNEPKLIESISDSLYDLKHONGLDDSHFISPE 887  
Db 707 KLAELKDDCVIRIEMNEKEDKINMLKEEY-----DKINTLKEQN-----E 750  
Qy 888 DISKTENGFRIRIPINKETGNSFIETEKIFSEYATHISKEISNIKDTIFDNVNGKLKVK 947  
Db 751 D-----KINTLKEQ--NEDKINTLKE--EY-----BHKINTMKEEYEHEKIN-----T 788  
Qy 948 VNLDAAEVNTLNSAFFIOSLIEYNTTKESLNSLVAM-----KVQVVAQIFSTGLNTI 1001  
Db 789 LNEQNEHKINTLNE-----QNEHKINTMKEEYEDQWNTLNEQNEKDNLSKEEYENKINQI 844  
Qy 1002 TDAS--KVVELASTALDEIDILPTLSEGLPIIATIDGVSILGAALKELSETNDPLLRQE 1059  
Db 845 NNNEIKIKDVNEEYEEVDKVLTLDEKKKQFDKEINYAHKAHEKE-----892  
Qy 1060 IEAKIGIMAVNLTAATAIIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATK 1119  
Db 893 -----QILLTEM-----EELKCO-RDNK 909  
Qy 1120 VID-YPKHISLAETAGFTLLDDKIIMPQDDVLVSEIDFNNSITLGKCEIWRAB--GGS 1176  
Db 910 YSDLYEKIKLKSICMIINIECCDDIENEDIIRIEEYINNKGK--EVEKEHKRHS 968  
Qy 1177 GHTLTDIDHFFSSPSITRKWLSIYD---VLNIKKEKIDFSKOLWLPNPNRVFGVE 1233  
Db 969 SPNLSKEKEKFPN-----SIEDKSHLKKKHEKDLSSKKEI--EEKNK-----1011  
Qy 1234 MGWTFGRSLDNDGTLLDRIR-----DHVEGQFYWRFAFTADALITLKLKPRY 1282  
Db 1012 -----KIKELANDIKKLQDEILVYKQSNAAQVDHKKKSWI-----LLKDKSKKI 1057  
Qy 1283 EDTNVRINDGNTRSPI-----VPVITTEQIR--KNLSYSFYGGSGYSLSLSPNMNID 1335  
Db 1058 KQENQINVEKNEEKDLKKDDDEIRILNEELVKYTI-----LYNLKKDPLQND 1108  
Qy 1336 -LNLVENDTWID---VDNVKNITIEDEIQG-----ELEENILS-----KLNTED 1379  
Db 1109 LUSKDINSLTNEGCMVDKIEHILDYDEEINKRSNLFQKNEICSTTVEWELNNKK 1168  
Qy 1380 NKIILNHTINFG-----DINESNRFISLTFISILEDINIIE-----I 1418  
Db 1169 NELIEENKLNLDVQGGKLLKXQVKKQKEIEKLNKQLTKCNQIDELNEVEKLNENI 1228  
Qy 1419 DLVSKS-----YKILLSGNCKLIENSDDIOQ---KIDH-----1449  
Db 1229 ELITYSNDLNNKPFMKENLNMKLDENEDNIKKMSKIDDMKEIKYREDEKKNLNEIN 1288  
Qy 1450 -----IGFNGEHQKY-----IFYSYIDNETKYNF-IDYSKK-----1480  
Db 1289 NLUKKNEDWCIKYENENIKYDICKVKEEMSLTYKETSLSKYBQIKVYDEKCSQVDEIRF 1348  
Qy 1481 ---EGLFTAFFSNE---SIIRNIYMPDSNNLFTYSSKDLKDRIINKGDVKKLIGNYFK-- 1533  
Db 1349 QYDEKCFQDEINKKYGALLNI---NITNKWDSKVDRNNELIISVDNKVEGIANYLKQI 1405  
Qy 1534 -----DDMKVISLFT---IEDTNTIKLNGVYLDENGVAQILKFM-----NNAKS 1574  
Db 1406 FELNEEIRLKGBEINKISLLYSNELNEKNSYDINMKHIOEQ-----LLPLEKTNKENERK 1460  
Qy 1575 ALNTNSLNN-----PLESINI-KNIFYNLDPN---IEFILDTFIISG 1615  
Db 1461 IINLTSQVSDAYKKSDSKLCAQFVDVNIYGNISNNIRTYENKYEBEMFDTNIEKN 1520  
Qy 1616 SNSIGOFELICDKDK--NTPQYFINKPIKET-----SYTLVGNRQNLIVE---PSYH 1663  
Db 1521 GHLSKYIHLEENPRCKMIYENENIKSSNKIIGLYNSRYGLREDLCKEEIVPS-- 1578  
Qy 1664 LDGSGNISSTVNFOSKYLYGIDRYNKKVII 1694  
Db 1579 --KIGNISKNENNNKKN-NTCDGYDEKVTI 1606

## RESULT 7

AAG82935  
ID AAG82935 standard; Protein; 5024 AA.  
XX AAG82935;  
AC AC  
DT 03-SEP-2001 (first entry)  
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2964.  
DE Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
KW vaccination; endocarditis.  
KW Staphylococcus epidermidis.  
OS Staphylococcus epidermidis.  
XX WO200134809-A2.  
PN 17-MAY-2001.  
PD 09-NOV-2000; 2000WO-US30782.  
XX 09-NOV-1999; 99US-0164258.  
PF (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
PI WPI; 2001-316495/33.  
XX N-PSDB; AAH53785.  
DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX Claim 18; Page 779-781; 2188pp; English.  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX Sequence 5024 AA;  
SQ  
Query Match 4.1%; Score 357.5; DB 22; Length 5024;  
Best Local Similarity 19.5%; Pred No. 1.4e-09;  
Matches 367; Conservative 310; Mismatches 631; Indels 573; Gaps 100;  
Qy 20 QDEYVATLNALEEYHNHNSSESVVEKYLLKLDINN-----LTDN-----YLN-- 61  
Db 2611 OKD---TILNHI--PSAPTRSQVGKIASAKQLNNTMKALRDSIADNNILQSSKYFNEB 2665  
Qy 62 TYKSGRNKALKKFEYLTMEVLELKNLSLTPV-----EKNLHFIIWIGG 105  
Db 2666 SEQQNAVQAVNKAKNII-----NQDPTVMADEIQSVLNEVKOTKDLH-----GD 2713  
Qy 106 Q--IND-----TAINYINQWQVNSDYTVKFYDSN-----AFLNLTAKT 144

Db 2714 QKLANDKTDAQTALNALYNQAGNLETKVQ---NSNSRDEVKVQVQLANQLANDAMKK 2770  
 Qy 145 IVESATNTLESFRENLDNPEFYKFKRMEIILYDKQKHFIDYKQIEB--NPEF-- 200  
 Db 2771 LODALTGN--DAIKOTSINIEDTSQ-----QVNFDE---YDRGKNIVAEQTPNPNMSP 2819  
 Qy 201 -----IIDNIITKLSNEYSKDEALNKYIEBSLNKITANNNDIRNLEKFADEDLVRLY 255  
 Db 2820 TWINTIADKI-----TEAKNDLHGQKUL--EQAQQOSINTINQMTGLNQAQKEQL---- 2867  
 Qy 256 NOELVERWNLAASDILRISMLKEBGGVYLDVILPGIOPDLFKSINKPDSITNTSWEMI 315  
 Db 2868 NOEIQOTQTRSEVHVQINKAQLNDS-----MNTLRQSITDEHEVKQTSNYIN 2915  
 Qy 316 KL-----EAIMKYEVI PGYTSKNFMDLDEEVBORSFESALSSKSKSEIFLPLDDIK 367  
 Db 2916 ETVGQATYNNVAVRQKQIINOTSNTMPL--EVERATSNVKTSK-DALHGERELNDNK 2972  
 Qy 368 VSPLEVKIAPANSV--INQALISLKDSYCSDL-----VINQIKNRYKIILN----- 411  
 Db 2973 NSK-----TFVNLHDLNQA--QKEALTHIEQATIVSQVNNIYNKAKALNNDMKKUL 3024  
 Qy 412 -----DNLPSINEGDTFTWKIPSPDKLASISNEDNMFM-----KITVYKVG 457  
 Db 3025 DIVAQDNVRSQNNYINEDSTPQMYND--TINHAQSIIDQVANTPMSHDIEINAIN-- 3079  
 Qy 458 PAPDVRSTIN-LSGPGVYTGAYQDMLMFKDNSTNHLPELNRNPEFP-KTKISOLTE-- 513  
 Db 3080 ---NIKHAINALDGE-----HKLQQAQKANA---NLLNSLNDLNAPORDAINLVNEA 3126  
 Qy 514 -----QBITSLMSFNQA-----RAKSQF---EYVKGYFEGALGEDDNL 549  
 Db 3127 QTRKEVAEQLOQAQALNDAMKHLRNSIQNSSVRQESKYINASDAKKEQYNHAREVENI 3186  
 Qy 550 DPAQNTVLQDVYSKILSSMKTRNKEYIHYIVQOGDKISYEASC----- 595  
 Db 3187 INEQHTPLDKEII-KQLTDAVNAQAND-LNGVELLDADQNAHQSIPTLMLHNOAQNAL 3244  
 Qy 596 -----NLFSDKDPYSILYQ-----KNIESEATYVVYVADAE-IKEIDKVRPYQIS 641  
 Db 3245 NEKINNNAVTRAKVAALIGQAKILDHAMEENLES-----IKDKEQVKQSSNY-----IN 3292  
 Qy 642 NKRNIKLF---IGHGKSEFNTDTFANLDVDSLSSEITILMLAKADISPKYIEINLGC 698  
 Db 3293 EPDVOQETTNNADVHTIELNQTVPNTLSIEDIEHAINEV-NOAKKQLRGK----- 3342  
 Qy 699 NMFYSIYAEETYPGKLLIKIDRVSELMPSISQDSITVSANQYEVINEEGKR----- 752  
 Db 3343 ----QKLYQTDIDLADKELSKLDLTSQSSSISNQIYTTAKTRTEVAQAIEKAKSLNHAMK 3398  
 Qy 753 -----EILDHSGKWINK-----ESTIKDISKEYISFNPKENKILVSKYLH 795  
 Db 3399 ALNKIYKNADKVLID-SSRFINEQDPEKEAYQQAINHVDSIHRTQNPENMDPTVINS-ITH 3456  
 Qy 796 ELSTLLQEIERN-----ANSDDIDLEKKVMLTECEIN-----VASNIDR- 834  
 Db 3457 ELETAGNNLHGDKLAHAKQDAANVINGLIHLNVAQREVMINTNTATREKVAKNLDNA 3516  
 Qy 835 QIVEGRIEPAKNLTSIDINYNKFEKLIBESIDSLYDLKHONGLDSDHFISEFISKTEN 894  
 Db 3517 QALDKAMETLQOVVAHK-NNILLNDSKYLN--EDSKYQQQVDRVIADAEOQLLQNTNPTLE 3573  
 Qy 895 GFRAIRINKETGNSIFIEFEKEIFSEYATHISKEISNIKDTIPDNVNGKLVKYNLDAAH 954  
 Db 3574 PYKVDIVKDNV-----LANEKILFG--AEKLSYDKSNANDEI-----K 3609  
 Qy 955 EVNTLNSAFFIQLSLEYNTTKBSLNSLVAMKVQVVAQLFST---GLNTITDASKVVELV 1011  
 Db 3610 HMYNLNA-----OKQSIKWMISHAALRTEVKQLLOQAKTLDKAMKSLDKTQVV-IT 3661  
 Qy 1012 STAL-----DETIDLLPTLSEGLPIIATII--DGVSLG---AAIKSELSETNDPL-- 1055

Db 3662 DTTLPNRYTEASEDKKEKVD--QTVSHAQAIIDIKINGSNVSLDQVRQALEQLTQASENLOG 3719  
 Qy 1056 -LRQIEIAKI-----GIMAVNLTAASTA--IVTSALGIASGFSILLVPLAGISAGIPIS 1105  
 Db 3720 DORVE-EAKVHANQTIIDOLTHLSLQOQTAKESVKNATKLB-----TAT 3763  
 Qy 1106 LVNNELIQQDKATKVIDYFKHISLAETEGAFLLDDKIIMPDDLVLSLSEIDPNNNSITIG 1165  
 Db 3764 ASNNALANLKVWGKLEQFINHADSIENSNDYRQADDDKIAYDDALEHGQDIQKSNATQN 3823  
 Qy 1166 KCEIWRAGGSGGHTLTDIDHP-FSSP-SITYRKPWLSIYDVNLNKK-----EKIDFSKDL 1219  
 Db 3824 EAKQALOQLINAEATSLNGFERLTHARPRALEYIK---SLEKINNAOKSALEDKVTQSHDL 3880  
 Qy 1220 MVLPNAPNRVFGYENGWTPGFRSLDNDGDKLLDRIRDHYEGQFYWRYFAFTADALITKLK 1279  
 Db 3881 LELEH-----LVNEGTLNDINGE-----LANAIVN--- 3906  
 Qy 1280 PRYEDTNVRIINLDGNTRSFIVPITTEQIRK-NLSYFSGGSGYSLSLSPYNNMIDNML 1338  
 Db 3907 -NYAPTASIN-----YINADNLKONFTQAINNARDALN---KTQCNLDENA 3951  
 Qy 1339 VENTDWTVIDVD-----NVVKNITIESDEIOKGLIENI--LSKLNI-EDNKIILNNHTI 1389  
 Db 3952 I--DTFKDDIFTKDALNGIERLTAAKSAAEK--LIDSLKFINKAQFTHANDEIMWNTNS- 4006  
 Qy 1390 NPYGDIENSNRPFISLTFISLEIDINIIEIDLVSYSKILLSGCMKLIENSDDIOQKIDH 1449  
 Db 4007 -----IAQLSRIVNQAFD-LNDAMKSLRDELNNQAFPVQASSN---YINSDEDLKQFDPH 4057  
 Qy 1450 IGFNGEHQKIFYSYIDNETKYNFYDKYKGLFTABFSNIESIIRNIYMPDSNNLFIYS 1509  
 Db 4058 A-----LSNARKV-----LAKENG----- 4071  
 Qy 1510 SKDLKDIDRIINKGDKVLLIGNFYKDDMKVLSFTIEDTNTKLVGYLDENGVA----- 1563  
 Db 4072 -KNLDEIQI-----EGLKQVIEDTKD-ALNGIQRSLKAKAKAIQYV 4110  
 Qy 1564 QILKEMNNAKALNTSN-----SLMNFLESINIKNIYNNLDNPTIEFIDTNTFIISG 1615  
 Db 4111 QSLSYINDAQRIHIAESNIHNSDDLSLANTLSKA-----SDLDNAMKDLRDT--LESN 4161  
 Qy 1616 SNSIQGFELICDKOKNIQPYF 1636  
 Db 4162 STSVSPSVNYINADKNLQIEP 4182  
 RESULT 8  
 ABP38314  
 ID ABP38314 standard; Protein; 10182 AA.  
 XX  
 AC ABP38314;  
 XX  
 XX 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX

PI Doucette-Stamm LA, Bush D;  
 XX WPI; 2002-381255/41.  
 DR N-PSDB; ABN90859.  
 XX  
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 XX Disclosure; SEQ ID 3159; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP3124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 10182 AA;  
 Query Match 4.0%; Score 348; DB 23; Length 10182;  
 Best Local Similarity 19.4%; Pred. No. 1.1e-08;  
 Matches 354; Conservative 317; Mismatches 663; Indels 494; Gaps 97;  
 QY 20 QEDYVAILNALLEEYHNMSESSVVEKYLKLDINN-----LTQN-----YLN-- 61  
 DB 7557 QKD---TILNHI--FSAPTRSQVGEKIASAKQLNNTMKALRDSIADNNEILOSKYFNE 7611  
 QY 62 TYKSGRNKALKKFKYLETWLEVLKNNSLTPV-----NDQTPWANDETQSULNEVKQTKDNLH-----EKNLHFTWIGG 105  
 DB 7612 SEQQNAVQAVKAKII-----NDQTPWANDETQSULNEVKQTKDNLH-----GD 7659  
 QY 106 Q---IND-----TAINYINQKOVNSDYTVKFVYDSN-----AFLINLTKT 144  
 DB 7660 QKLANDKTDQAATNALNLNQAQRGNLETKVQ---NSNSRPEVQVQVQLANQLNDAMCK 7716  
 QY 145 IVESATNTLESFRENLDPEFYNKFKYKRMELIYVDKQKHIDYIKSQIER--NPEF-- 200  
 DB 7717 LDDALFQ---DAIKQTSYINEDTSQ-----QVNFDE---YTDGRKNIVASQTPNMPSP 7765  
 QY 201 ----TIDNLIKTYLSNEYSKOLEALNKY---IEESLNKITANNQNDIRNLEKFADEDLV 252  
 DB 7766 TWINTIADKI-----TEAKNDLHGQVKLQKQKQKQSQSINTI-----NQMTGLNQAQKEQL- 7813  
 QY 253 RLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPFIQDLPFKSINKPDSITNTSW 312  
 DB 7814 ---NQBIQQTQTRSEVHVQVINKAQLNDS-----MNTLRQSIITDEHEVKQISN 7858  
 QY 313 EMKL-----EALMKYKEIPGVTSKNFMDLDEEVORSFESALSSKSDKSEIFLPJD 364  
 DB 7859 YINETVGNQYANNVADVQKLIQINQTSNPTMPL--EVERATSNVKISK--DALHGERELN 7915  
 QY 365 DIKVSPLVKIAFANNSV--INQALISLSDSYCSDL-----VINQIKNRYKILN---- 411  
 DB 7916 DNKSK-----TFAVNHLNLAQ-----QKEATHEIQATVISOVNNIYKAKALNDMK 7967  
 QY 412 -----DNLNPSINEGTFDNTMTKIFSDKLASINEDNMFMFI-----KITNYL 454  
 DB 7968 KLKDIVAQDQNVQRQSNYINEDSTPQMYND---TINHAQSIIDQVANPTMSHDEIENAI 8024  
 QY 455 KVGFPADVRSTIN--LSCPGVYTGAVQDMLLMFKDNSTNHLHPELNFEPF--KTKISQIT 512  
 DB 8025 N-----NIKHAINALDGE-----HKLQAQKANA---NLLINSNLNDLAPORDAIRNLV 8069  
 QY 513 E-----QBITLSWSPNQ-----RAKSRF---EEYKGVFEGALGED 546  
 DB 8070 NEAQTREKVAEQLSQAQLNDAMKHLRNSIQNQSSVRQESKYINASDAKQYINHAVREV 8129  
 QY 547 DNLDFQANTVLDKDYVSKKILSSMKTRNKEYIHYIQLQGDKISYEASCN--LPSKDPYS 604

DB 8130 ENINEQHPTLDKEII-KQLTDGVNQAND--LNGVELLDADQNAHQSIPTLMLHNAQAO 8187  
 QY 605 SILYOK-NIEGSETAYYYVADAETKEIDKYRPIQVSNRNIKULTFIGHGSEFNTDTF 663  
 DB 8188 NALNEKINNNAVTRTEVAALIGQAKLLDHAMENLEESIKDEQVK-----QSSNYINEDSD 8242  
 QY 664 ANLVDLSLSEIETILN-----LAKADISPKYIEN-----LLGCNMFYSYIAEETYP 712  
 DB 8243 VOETVDNAVHVTIELNQTVPNTLSIEDIEHAINVNOAKQLRG-----KQKLYOTIDLA 8298  
 QY 713 GKLLLIKIDRVSELMPISQSDSITVSANGYEVINEEGKR-----EILD 756  
 DB 8299 DKELSKLDDLTSQOSSSISNOIYTAKTREVAQAETAKSLNHAMKALNKVYNADKVL 8358  
 QY 757 HSGKWINKB-----ESIHKDISKEYISFNPKNKIIVKSKYLHELSTLQEIERN-- 807  
 DB 8359 -SSRFINEQPEKAYQQAINHVDSIHRTQNPMDPTVINS-ITHELETAQNHLHGDQK 8416  
 QY 808 -----ANSSDIDLEKKVWLTECEIN-----VASNIDR-QIVEGREIEAKNLT 848  
 DB 8417 LAHAQOAAANVINGLIHLNVAQREVMINTNTATTREKVAKLDNAQALDKAMETLQQV 8476  
 QY 849 SDSINYIKNEFKLIESISDSLYDLKHQGLDSDSHFISFEDISKTENGPRIRINKETGNS 908  
 DB 8477 AHK--NNILNDSKYLN--EDSKYQOQYDRIADAEOQLLQNTTNTPTLEPKYVDIVKDNV--- 8530  
 QY 909 FIETEKIFSEYATHISKEISNIKDTTFDNNVNGKLVKKNLDAHAHVNTLNSAFFIOSL 968  
 DB 8531 --LANEKILFG--AEKLSYDKSNANDEI-----KHMNYLNA----- 8563  
 QY 969 IEYNTTKESLNSLVAMKVO--VYAFSTGLTNTITDASKVVELVSTAL----- 1015  
 DB 8564 -QKQSIKMI SHAURTEVKKLQAKILDEAMKSLSDTKQV--ITDITLPNTEASEDK 8621  
 QY 1016 DETIDLLPTSEGLPIIATII--DGVSLG--AAIKELSETNDPL---LRQIEAKI--- 1064  
 DB 8622 KEKVD--QTVSHAQAIIDKINGSNVSLQVQALQOLQASENLDQDQVE-EAKVHANQ 8678  
 QY 1065 ---GIMAVNLTAASTA--IVTSALGIAGSFSILLVPLAGISAGISLVNNEILIQDKATK 1119  
 DB 8679 TIDQLTHLSLQOQTAKESVKNATKLEE-----IATVSNAQALNKWVGK 8723  
 QY 1120 VIDYFKHISLAETEGAFLLDDKIIIMPODDLVLSEIDFNNSNITLGKCEIWRPAGSGSHT 1179  
 DB 8724 LEQFINHADSVENSNYRQADDDKIIAYDEALEHGQDIQKTNATQNETK----- 8772  
 QY 1180 LTDDIDHFFSPSITYRKPWLSIYDVLN--IKKEKIDFSKDLMLVLPNPNRVFGYEMGWT 1238  
 DB 8773 -----QALQOLYIAYETSLNGFERLNHARPRALEYIKSLEKINNQAQSALEDKVTOSH 8824  
 QY 1239 GFRSLD---NGTKLLDRIRDHYEGQFYWRVAFIADALITKLKPRYEDTNNRINLDGNT 1295  
 DB 8825 DLLELEHIVNEGNTLNDIMGE-----LANAIVN---NYAPTAKASIN----- 8862  
 QY 1296 RSFIVPVITTEQIRK-NLSYSFYGGSGYSLSLSPYNNNIDNLVENDTWIDVD----- 1349  
 DB 8863 -----YNAONLRKDNFTQAINNARDALN---KTQGNQLDFNAI--DTFKDDIDFKTKDA 8911  
 QY 1350 -NVKNITIDESIQKGLIENI--LSKLNI--ENKIIILNHTINFYGDINESNRFSISLT 1405  
 DB 8912 LNGIERLTAAKSAAEK--LIDSLKFINKAQFTHANDEIMTNS-----IAQLSRIVNQA 8963  
 QY 1406 FSILEDINIITIDLVSKYKILLSGNCMKLIENSDDIOQKIDHIGFNGEHQKIFYYSYI 1465  
 DB 8964 FD-LNDAMKSLRDELNNAQFPVQASSN---YINSDLEDLQKQFDHALSNA--RKVL----- 9012  
 QY 1466 DNETKYNG-FIDYSKKEGLFTAESNESIIRNIYMPDSNNLFIYSSKD-LKDIRIINKGD 1523  
 DB 9013 ---AKENGNLDEKIQIQL-----KOVIEDTKDALNGIQRLSKAK 9049  
 QY 1524 VKLLAIGNYKODMKVYSLSFTIEDTNTIKLNGVYLDENGVAQILKFMNNAKSALNTSNSIM 1583





Db 1035 -NDMNNLIDMNEAIIETVN-----GVINNIIIDRKDNNSRDKMEKEMEKEMKQM 1084  
Qy 1057 RQIEAKIGIMAVNLTAATAIVTSALGTASFSILLVPLAGISAGIPSLVNNELIQDK 1116  
Db 1085 EKEMEK---VMEKEMEKMEKEVEKEL-----KNEMNNRNNRNNRNNRNNRNN 1129  
Qy 1117 ATK---VIDYFKHISLAETEGFTLLDDKIIMPDDVLSEIDFN--NNSITLGRKEIWR 1171  
Db 1130 IYKNEIYVNDNDKEIIVNEE-----KKLIYFPN--YESDVHRKNNNSININNCK--- 1177  
Qy 1172 AEGSGHTLTDODTHFFSSPITYRKPMLSIYDVNLKKEKIDFSDKMLVLPNAPNRVFG 1231  
Db 1178 -----DDYN-----NILKEYVDS-----CLAQKEENI-- 1200  
Qy 1232 YEMGWTPGRSLDNDGTCLLDRIDHYEQFWRYFAP-----IDALIT 1276  
Db 1201 -----FRPLFNKK--DKV-----WKRFNKNNTIITHNEMKRIYQTINK 1241  
Qy 1277 KLPK-----RYED---TNVRINLDGNTRSFIVPVTTEOIRKNLSYFSGSGSYLSLS 1328  
Db 1242 NVPPIYFNRYENFLNHLTYNPPKNDLPKLSYKVMNNIR-NLIYANKHINNNDYMK 1300  
Qy 1329 PYNMNI-----DLNLVENDTWI-----DVDNVVKNITIE----- 1358  
Db 1301 LYNQNIYTLKYQVANIDNDHHICKGGGLDYINMNSICEKNRDKTYLNKIFHYKKKD 1360  
Qy 1359 -----SDETKQBELIENILSKLNIEDNKILNNH--TINFYGD-----INESNRFISL 1404  
Db 1361 ARPFINDEIGSNIDYDIKKYSNDENNYKLAENKMNISMNDEDMITPLNSEHGNFPSC 1420  
Qy 1405 TFSILEDINIITHIDIVSKYKILLGNCMKLIENSSDIQOKIDHIGFNGEHQKIFYSY 1464  
Db 1421 QPNLEKKSTYIDNLYDSNMDFTTEKYNFVNNENDL-----FNTKRWKFNFSK- 1471  
Qy 1465 IDNETYNGFIDYSKEGFTABFNSIESIRNIYMPDSNNLFYSSKOLKDIRINKGV 1524  
Db 1472 -GKNLFNNKFFVNSNEDGVFSF-FKNMNLREL-NKSNLSKLESYKSNNSNCKGDD 1528  
Qy 1525 KLLIGNY-----FKDDMKVSL-----SFTIEDNTI--KUNG---VYLDENGA 1563  
Db 1529 N--IGNMENNNTTNVTIASDEHISTKGDIDHESFRDDNDCLILLKTEGRSKYSD----- 1581  
Qy 1564 QILKFNNAKSALNTSNLWFLSINIKNIFVNNLDPNIEFLDTNFIISGS-NSIGOF 1622  
Db 1582 --ITLNEKSNLNDNETINEYEV-----CSNIDVN-----EWEDKVGTCNSVGDK 1628  
Qy 1623 ELICDKDKN 1631  
Db 1629 ETEKNEKN 1637

RESULT 10  
AAB18222  
ID AAB18222 standard; Protein; 2496 AA.

XX AAB18222;  
AC AAB18222;  
XX 07-NOV-2000 (first entry)  
DT  
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:79.  
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX Plasmodium falciparum.  
OS  
XX WO200025728-A2.  
PN  
XX 11-MAY-2000.  
XX  
XX 05-NOV-1999; 99WO-US26796.

PR 05-NOV-1998; 98US-0107131.  
XX (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX  
PT Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
PS Disclosure; Page 185-192; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAY70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 2496 AA;

Query Match 3.6%; Score 314.5; DB 21; Length 2496;  
Best Local Similarity 18.8%; Pred. No. 1e-07;  
Matches 370; Conservative 316; Mismatches 730; Indels 549; Gaps 96;  
Qy 57 DNYLTYKSGRKNALKFKE---VLTMEVLELKN--NSLTPVEKNLHPIWGGINDTA 111  
Db 24 DNKSVLRSTKTKKKKICKSTFYVHEESEIKSWLRNSNRDRKGFPIFERLIKERK 83  
Qy 112 INVINQWKVNSDYTVKFVDSNAFLINTLTKTIVESATNTNTLESFRENLDPEFDYKPF 171  
Db 84 YICVNKYRNNK---LKWIYK-----NTYEKT-----KNICO---DYNL 117  
Qy 172 YRKRMEIIVDKQH-FIDYKSGQIBENPEFI---IDNIKITLSNEYSKDLBALNKYIE 227  
Db 118 PKCIGIIVDKNKEFTFFETFPENFDNIFYNNKYIFNIY--YMFEDYTKKKVKKIEGIE 175  
Qy 228 SLNKITANGNDIRNLEK---PAEDDLVRLYNQELVERWNLAASDILRISMLKEDGGY 284  
Db 176 NMNIRHNNNNNIFYVHKFFLENDDEKKKRDDI--KINIKLHNTRKLSVSE----- 228  
Qy 285 LQVDILPGIOP-----DLF-----KSNKPDSTINTSWEMIKLEAIKYEYIPG 329  
Db 229 -NVEUKPIYKQGERNETVNVLYEYTGGVKRSNNNNEIVVTSTEFHRIIVICFPTVKH 287  
Qy 330 ---YTSKNFMDLDEEVQVSFESALSSKDSKSIIFLDDIKVSPLEVKIAFANNVINO 386  
Db 288 SHIITSPH-DALNHIVE-----ENDKIKLSEIYSIPFYIYGNLGLKN--VITTG 335  
Qy 387 LISLKDSCYSDLVIN-----QIKNRYKILNDN---LNPSINEGTDPN 425  
Db 336 IVEFMPYPSRTQNNFTVTGANGEMNDLYKFDLKIIRIRIPRNTKKILGLSTNE----- 390



CC vaccines against *P. falciparum* infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with *P. falciparum*. Furthermore,  
CC (I) especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent *P. falciparum*  
CC infection, or they can be used to identify drug resistance in  
CC *P. falciparum*. Sequencing of the plasmidium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasite lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a pressing need for vaccines and new  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX

SQ Sequence 3973 AA;

Query Match 3.6%; Score 313; DB 21; Length 3973;  
Best Local Similarity 18.5%; Pred. No. 2.2e-07;  
Matches 375; Conservative 335; Mismatches 708; Indels 611; Gaps 90;

QY 11 KVVYV--KFRIOED-----EYVAIL-----NALE---EYNNSESSVVEKYLK 49  
DB 808 KIYILRKQNKNDTIYIYIEYYILLDYFKYFYNNMDNFYKNNMKSTVRKRYKN 867  
QY 50 KDINLTDNLYNTYKSGRKNKALKPKFYLTWVLELKNKSLTPVEKNLHFIWIGQIND 109  
DB 868 KNIND-----TKGNK-----NNNN-----NNNIY-----GDGDD 892  
QY 110 TAINYNQKDV-----NSDYTKFYVDSNAFL-----INTLKTIVESATN 151  
DB 893 NNI-YCHDDDDIYCHDDDDIYCHDDDDIYCHDDDDIYCHDDDDIYCHDDDDIYCH 951  
QY 152 NTLESRENLDPEPDYKPYRKRMEIYDKQKHFD-----YKSOJEEENPEFTI--- 202  
DB 952 -----YFK-----IPYDKNTFINLTYLFFPSLKVQBFSSRQIMIV 989  
QY 203 -----DNIK-----TYLSN-----EYS----- 215  
DB 990 RCLSPFLKKKINKLNAYIFQLFSYLENDQININEKGMVHRKSSKYHRRNQEEYSHNKT 1049  
QY 216 -----KDEALNKYIEESLNK-----ITANGNDIRNLEKFADELDVLYNQELVE 261  
DB 1050 NDNSVSNLYRDIE--NEYDENHLERRKDRNVFSSNNMDDKYNLSDFKYTK-ENMDIKE 1106  
QY 262 RWNLAASDIIRISMLKEDGGVYLDVILPGIOPDLFKSINKPDSITNTSWEMIKLEATM 321  
DB 1107 NFRIDIS--FLKIFFLNDVR-----QINLESNGRKDLKESKAKRRIQKLDVH 1153  
QY 322 KY-----KEYPGYTSKNFMDLBEVORSFESALSSKSDKSEIFPLDDIKVSPLEVY 374  
DB 1154 RYTENNNDKNYDGNFTF--LSQDEKSKSPDSDSCSDEKE----- 1195  
QY 375 IAFANNSVINQALISKDYCSDLVINQIKRYKILNDNLNPSINEGTDFNTMTKIFSDK 434  
DB 1196 -----SSKGLYGNDFVNSDHNNSNNSSNNSSNNSSNNSSNNSSNNSSNNSS 1241  
QY 435 LASISNEDNMFMK-----ITNVLKVGFPADVSTIN-----LSGPGVYTCAYOD 480  
DB 1242 VVVDPPDYNYECKDSNKGVVVNYF-YAHLFPFEKSYNNYVVEDISFDDIFILISMD 1300  
QY 481 LLMFKDNSTNIHLLEPELRFNFFPKTK-----ISQTEQEIITSLMSFNQARAK 528  
DB 1301 LWETNNNNLLNLNDLLKTYIEEKKKIYICTSLLLKIFKRIKKKSSYFLFYKA- 1359  
QY 529 SQFEYKGYFEGALGEDNDLPAQNTVLDKDYVSKKLSMKTNRKEYIHVIQLQGDK 588  
DB 1360 -----FENDI-----KLILDSINILIKWVW----- 1381

QY 589 ISYBASCNLFSDKPYSSILYQKNIIEGSETAYVYVADAEIKEDKIRIPYQISNKRNIKL 648  
DB 1382 -TFKNCNDFNRE--KNINIKLVKLFISYKLVKNYFLQIYYHFFYNNQIYNERKNYF 1438  
QY 649 -TFIGHGSEFNTDTFANL--DVSJSSEIETILNLAADISP-----KYI 691  
DB 1439 DNFFSIFSKYINKIFVEIYSSSSSTSSSFSFVNFVSKFYMMKMCISINNMIGWKYI 1498  
QY 692 EIN-----LLGCMESYSIYAEETYPG-----KLLAKIKORVSELMPSI 730  
DB 1499 NLERVKQVYBEHINMDVMHSHLHDIDVYGHNSYNNIYQKIISKYRGEEKDTLDVI 1558  
QY 731 SODSITVSANQVEVR-----INEEGRBILDSHGKWINKEESIIKDISKEYISFNP 782  
DB 1559 NTESVHQNRNEDDIDGINSLSLDVFNEMRNIIINH-----SLIKD----- 1599  
QY 783 KENKIIVSKYLH--ELST--LLOIRNANS-----SIDLEKKVMLTECEINVASNDR 834  
DB 1600 -HNDMCTKKRINIFQISSPATSEQLMNNHYTMNYLTDVMLLQKDYIYNDNM--NEHK 1656  
QY 835 QIVEGRIBEAKNLTSDSINYIKNEFKLIESDSLY--DLKHQGLD--DSHFISFEDIS 890  
DB 1657 QNVFNKPFDDNN 1716  
QY 891 KTENGRIRFRINKETGNSIFITEKEIFEYATHISKEISNIKOTIFONVNGKLVKKNL 950  
DB 1717 NEENELIQ--NNSMSSSIYI--DKLMD-----TKEMEPLFNKTKDMKNYNEOKNE 1767  
QY 951 DAAHEVNLT--NSAFFIO-----SLIEYNTTKESLSN--LSVAMKVQVYQAIFSTG- 997  
DB 1768 LISYPYNNMLQNNIIFVKFFLYVTQNLQIIFQNNYIFFLSDFLFNYKKKEYIEBKGN 1827  
QY 998 --LNTITDASKVVELVSTALDETDLPLSEGLPIATIIDGVSGLGAIKELSETNPL 1055  
DB 1828 QNVINIKDEDKHI-----TNIKDGDKHTNIKDGDKNTNIKDDDKXNTNM 1873  
QY 1056 LRQETEAIGIGAMVNLTAATAIVTSALGIASGFSILLVPLAGISAGISPLVNNELIQQ 1115  
DB 1874 KKNNKNVLTILMYSQESFY-----YSI-----FNTLINDYFLY 1911  
QY 1116 KATKVIDYFKHISIAETEGATFLD-----DKIIMPQDDLVUSE--- 1154  
DB 1912 K-----DYKSCFLYESLNTFFKNNFNFIYFLCKYSSGYLPLERIIKFMDFVIFCHFTK 1966  
QY 1155 -IDENNSITLCKEHWAEAGSGHTLDDIDHFPSSPSITVRKPLSIVLVNLKKEKI 1213  
DB 1967 FINENINDELVELLEVLYNGKCYELLRFL--FF-----FIKONDDITINIKY 2014  
QY 1214 DFKDMLVLPNAPNRFVGY--EMGWTGFRSLDNDGTKLLDRIDHYEGQFYRYFAP 1270  
DB 2015 IFDIIMCILEQYIAHVNYIYKKAWDVFFNKL-----KILN--LSLHFVNSIYFIFCDD 2068  
QY 1271 ADALITK-----LKPREDYTVNRINLDGNTSRFIVPVITEQIRKNLSYS---- 1315  
DB 2069 INAEIKRENDNNSKNDNN 2128  
QY 1316 FY-----GSGGSY-----SLSLSPYNNIIDLNLVENDTWDVDN 1350  
DB 2129 LFYNNVINKLHVNSINCLKNKLGYSKDEEPPKELNCKRYFLNY-----NKDFKKEIY 2182  
QY 1351 VYKNTIDESDEIQKELIENILSKINIEDNKIILNHNTHINFYGDINENRISLTFSTILE 1410  
DB 2183 VLYNLIAS-----EIFELIKAIYINETKIY--PLINICYDRNISINFINIDY--D 2231  
QY 1411 DINIIE-IDLVSKSYKILLSGNCKMLIENSDDIOOKI-----DHGFNGSH---QKYI 1460  
DB 2232 NLNSILEKYTYLHKKHOKHINKLVLLCKKSIHMHKVIYSIDDDHLLNNMLHLRRKNI 2291  
QY 1461 FYSIDNETKNGFTDYKKEGLFTAEPNSESIIRNIYMPDPSNNLFIYS-----KDL 1513  
DB 2292 YKYVLYNINEYNNFLDNHKKR--KRKFINYNNIOSSY---NNNNYINNTNTNFYEYHDY 2346  
QY 1514 KDRIIRKGDVKLLIGNYFKDDMKVSLSFTEDTTNTIKNGV--YL-DENGVAQILKFMN 1570

Db	2347	I A I K N I L H K K I E L D D Y ----- I C S R I L D T Q S O K T Y G E K N Y L F D V K N Y T Y N M N F I N	2398
Qy	1571	N A K S A L N T S N L M N ----- F L E S I ----- N I K ----- N I F N N L	1599
Db	2399	N N Y Q E N S I N D V I N G K K K M F T L Q I S E Y D K H T N Y N S F M D C V Q N H N I K O N S T N N M N H I	2458
Qy	1600	D N I E F I L D T N F I ----- I S G S N S I G Q E L I C D K D K N I O P P I N F K I K E T S Y T L V G N	1652
Db	2459	N T N N Y L H N F I S N Y N S F N V H D N K K I Y S N E N C K S D E I M Q K K - I D M S I W K N I D S I P P E T	2517
Qy	1653	R Q N L I V E P S Y H L D -- D S G N I S S T W I N F S Q K V L Y - G I D R Y N K V I I A P N L	1698
Db	2518	F I D S D K O P A N Y F O P I N S I N G S R S N N E K K K I Q I D N P V K E C E L L I N I	2566

RESULT 12  
AAE00425  
ID AAE00425 standard: Protein; 2184 AA.

DT 19-JUN-2001 (first entry)

DE P. falciparum telomerase reverse transcriptase full length protein.

XX Telomerase reverse transcriptase; TERT; ever shorter telomere;  
KW EST; therapy; stomach cancer; malaria; vaginal candidiasis.

OS Plasmodium falciparum.

FH	Key	Location/Qualifiers
23		

FT	Misc-difference	330	/label= Leu, Ile	/note= "Encoded by MTA"
FT				
FT				

```

11  /note= "Encoded by MIA"
FT  Misc-difference 335
FT  /label= Asp, Gly
FT  /note= "Encoded by GWT"

```

XX  
PN  
WO200127287-A2.

XX  
PD  
19-APR-2001.

10-OCT-2000: 2000WO-US27825.

XX 13-OCT-1999: 99US-0417485.

XX  
PA (RERE-) RES & DEV INST INC.XX  
PI Long DM. Love R:XX  
DR WPT: 2001-266411/27

DR WPI; 2001-266411/21.  
DR N-PSDB: AAD03729.

xx Telomerase reverse transcriptase genes and proteins from *Plasmodium falciparum* and *Candida albicans* are used to detect infections of these pathogens in mammals -

PS Claim 10: Page 104-111: 138pp: English.

The invention relates to identification and use of telomerase reverse transcriptase (TERT) genes and proteins isolated from *Plasmodium falciparum*, *Candida albicans* and rice. TERT gene is also called as ever shorter telomere (EST) genes. TERT genes are used to detect *P. falciparum* and *C. albicans* infection in mammals. They are also used to diagnose the state of an infection in a patient and the relative amount of the pathogen in a cell, tissue, organ or organism. The potential binding partners and modulators of the activity of TERT genes and proteins are used to inhibit or promote growth of the pathogens and hence are used to treat e.g. stomach cancer, malaria and vaginal candidiasis in mammals.

CC The present sequence is *Plasmodium falciparum* TERT full length

[illegible]

Db 1436 NNNKFMCTNHEQDTE---EKGNTQNKERKHDIVIGPIYNNSPDSTTTTHSSNNYKGNH 1492  
Qy 789 VKSKY-----LHELSTLLOE-----IRNANSSDIDLEKKVMLTECEINVASNIDROI 836  
Db 1493 VSGDYKNDGLLKHGNNMNECYVKOTKCNVNNNNNNNNNNNNNNNNNNNNNNNNKNCVTNNKND 1552  
Qy 837 VEGRIEAKNLTSDSI---NYIKNEPK--LIESISDLYDLKHQGLDD---SHFISPD 888  
Db 1553 I---IKYHTIDTNSKNHTYFNKFLNDFDKLISNIYGLPQGSLSNLSLYAYLD 1609  
Qy 889 ISKTENGPRIRFINKTGNSIFIEETEKEIFSEYATHISKEISNIKIDTIFDNVNGK-----943  
Db 1610 --KNEEFQNLVYSEKQNNKYFLANGT---CNVFNLSLILRFIDDFLITLKNKNIKIF 1664  
Qy 944 ----LVKKNLDRAEVNTLNSAFFIOSLI-----EYNTKESLSNLSVAMKV 987  
Db 1665 KNLLKKKI---WGSNINSKTKIKPIIYKNDLLIYNFONKYQKKYKIKRKKKIQSV 1721  
Qy 988 ---QVYAOQLFSTGLNTITDASKVVELVSTALDETIDL--PTLSEGLPIIA--TIIDGVSL 1041  
Db 1722 RKRKHQNLVANKKHTSVQK-----DKINKYINLIHPTIQKNDVLSNNSIMNPERI 1775  
Qy 1042 GAAIKELSETNDFLLRQREIAKIGIMAVNLTAATAVTSALGIASGFSILLVPLAGISA 1101  
Db 1776 --YIKESHKS-----SSIRT 1789  
Qy 1102 GIP-SLVNNELIQDKATKVIDYFKHISLAETEGAFITLDDKIIMPQDVLVSEDFNN 1160  
Db 1790 DIPNSVNDIDIEYNQSD-----NNSYSTNNLYNNINMTQNG-----DNN 1829  
Qy 1161 SITLGCETWRAEGSGHTLTDDIDHPFSSPSITYRKPLWSYDVNLNIKEIKIDFSKDL 1220  
Db 1830 NVNIFK-----HVQNSFOCFNSN-----NLYEKIDKENNI-----1861  
Qy 1221 VLPNAPNRVFGEMGTWPGFRSLDNDGTLLDRIDHYEGQFYRWYFAPFADALITKLP 1280  
Db 1862 ---SQINRKLCKSRNFTKSRKINTLTYLQDKV-----IKILCKK 1900  
Qy 1281 RYEDTNVRNLGNTSRFIVPVITTEQIRKNLSYFYGGSGYSLSLSPYNNI--DLNV 1339  
Db 1901 KYIKHIKKMKYNNFQNF-----KKLKKQFP-----HNASFELKINKINKIRLNKL 1949  
Qy 1340 ENDTWIDVDNVKNITIESDEIQKGELIENTILSKLINIEDNKILNHTINFGVDINES 1399  
Db 1950 KKR-----KNHSI-----NTPVTSIE---WLNNSYTFDF---INNS- 1980  
Qy 1400 RFISLTFSEIDENIIIEIDLVSYSKYILLSGNCKMLIENSDDIOOKIDHIGFNGEHOKY 1459  
Db 1981 -----IQSTSYF--WKNKCDATIRNHLHNV-----IDKNKTY 2014  
Qy 1460 IFVSYDNTKNGFYDYSKKEGLFTAEPFSNESIIRNIYMPDSNLIPIYSSKDLDIRI 1519  
Db 2015 FMKNLVENRIVRNIISKQKQCSLYKNK-----QNVYFCYKNFSLKSLKFCICI 2067  
Qy 1520 NKGDVLLGNVFKDPMKYSLSFIEDTNTIKLNGVYLDENGAQILKPMNAKSALNTS 1579  
Db 2068 -----KTLKMPNAFT-----NSTY-----NTKFL-----2088  
Qy 1580 NSLMNPLESINIKNIYNNLDPNIEFILTNTFTISGNSIGOFELICDKDKNIQPYFINF 1639  
Db 2089 -FLISYNNKMLIKN-----KKLKFV--KLFLIQTAEAPRYARIFNOODSFPYCLQHF 2138  
Qy 1640 -KIKETSYLY-VGNRQNLIVPSYHLDDSGNISSTVINFSQKLYIGIDRYNKKV 1692  
Db 2139 RKIKKRLINKYKIGHKNLLREFF-----LNFNIKELYNWPYMFKI 2182

RESULT 13  
AAW54145  
ID AAW54145 standard; Protein; 1639 AA.  
XX  
AC AAW54145;

XX 25-MAR-2003 (updated)  
DT 23-SEP-1998 (first entry)  
XX P. falciparum synthetic gpl90 protein.  
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;  
monoclonal antibody; passive immunisation; parasite.  
OS Plasmodium falciparum.  
Synthetic.  
XX WO9814583-A2.  
XX 09-APR-1998.  
XX 02-OCT-1997; 97WO-EP05441.  
XX 02-OCT-1996; 96DE-1040817.  
XX (BUJA/) BUJARD H.  
PI Bujard H, Pan W, Tolle R;  
DR WPI; 1998-240088/21.  
XX N-PSDB; AAV21451, AAV35363.  
XX Recombinant production of complete gpl90/MSP-1 Plasmodium surface  
protein - useful in anti-malaria vaccines, also stabilising genes by  
reducing their AT content  
XX Example 1; Fig 3c; 48pp; German.  
XX This sequence represents a modified Plasmodium falciparum gpl90/MSP-1  
(merozoite surface) protein. The gene encoding this protein has been  
stabilised by reducing the AT content of the nucleotide sequence. Such a  
protein is useful in vaccines against malaria or for producing monoclonal  
antibodies (for passive immunisation). The complete gpl90 protein can now  
be produced outside the parasite and has, at least over extended regions,  
the native pattern of folding. Larger amounts of the protein can be  
produced recombinantly than would be possible using the parasites as  
source.  
XX (Updated on 25-MAR-2003 to correct PR field.)  
XX Sequence 1639 AA;  
SQ  
Query Match 3.6%; Score 309; DB 19; Length 1639;  
Best Local Similarity 19.5%; Pred. NO. 1.1e-07;  
Matches 351; Conservative 280; Mismatches 554; Indels 618; Gaps 94;  
Qy 36 NMSESVVEKYLKLDINNLTNDYNTYKSGRNKALKKFKLEYLTWEVLEKNNSLTPVE 95  
Db 112 NPSDSSSDAKSYADLKHRVRYLLT-----IKELYPQLFDLTNHLMLTCLD 159  
Qy 96 KNLHFIWIGQINDTAINY-INWKDVNSDYTVKPYVDSNAPLINTLTKTIVESATNNTL 154  
Db 160 -NIH-----GFKLIDGYEIN-----ELLYKLN-FYDLLRAKLNDCANDVC 201  
Qy 155 E-SFRENLDPEFDYNNK-----FYKRMBEIIYDKQKHFDIYKQSOIEENPEFIIDNIKIT 209  
Db 202 QIPFNLKIRANELDVLKLVFGYRKLPNIDIKNVGMEDYIKKN-----245  
Qy 210 LSNEYSKOLEALNKYIEESLNKITANNNDIRNLEK-----FADELVLRYNQELVERNVL 265  
Db 246 -----KKTIENTINELIEES-KKTIDKNKNATKEEEKKLYQAOYDL-SIYNQOLBEAHL 298  
Qy 266 AASDILIRISMLKEDGGVYLDVILPGIQ-PDLFKSINKPSITNTSWEMIKLEAIMKYK 324  
Db 299 ISVLE-KRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNL-----LDKNKKIE 348  
Qy 325 EYIPGYTSKNFMDLDEEVQVSFESALSKSDKSEIFLPLDDIKVSPLEVK--IAFANNSV 382  
Db 349 EH-----EKEI-----KETAKTIKFNIDSFTDPLELEYVLRKNKI 386







KW Plasmodium falciparum; Chromosome 2; human malaria parasite; vaccine;  
 XX antimalarial; malaria; protozoicide; infection; insecticide.  
 XX Plasmodium falciparum.  
 XX WO200025728-A2.  
 XX 11-MAY-2000.  
 XX 05-NOV-1999; 99WO-US26796.  
 XX 05-NOV-1998; 98US-0107131.  
 XX (HOFF/) HOFFMAN S.  
 XX (CARU/) CARUCCI D.  
 XX (GARD/) GARDNER M.  
 XX (VENT/) VENTER J C.  
 XX Hoffman S, Carucci D, Gardner M, Venter JC;  
 XX WPI; 2000-365347/31.  
 XX Proteins encoded by chromosome 2 of the human malarial parasite,  
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P.falciparum infection -  
 XX Disclosure; Page 120-124; 577pp; English.  
 XX The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
 CC vaccines against P. falciparum infection comprising (i) or (ii).  
 CC (i) and (ii) are useful for the development of vaccines against  
 CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal  
 CC antibody, raised to immunogens comprising the sequences of (i), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (i) (especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
 CC subsequent identification of proteins encoded by it will help to expand  
 CC our understanding of parasite biology, a process hampered by the  
 CC complexity of the parasitic lifecycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many  
 CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.  
 XX Sequence 1516 AA;  
 SQ  
 Query Match 3.5%; Score 306.5; DB 21; Length 1516;  
 Best Local Similarity 18.2%; Pred. No. 1.4e-07;  
 Matches 311; Conservative 278; Mismatches 494; Indels 627; Gaps 81;  
 QY 6 KAOLQVYVYKRIQEDVAVTALNALEEVHNMSSSVVEKYLKLDINNLT-----NYL 60  
 DB 131 KTAELKVTNYQ-----RTLNSMKSNNKNDNS-----NNIEDKTPNKT 173  
 QY 61 NTRYKSG-----RNKALKKFEYL-----TWLE 85  
 DB 174 NTQNSKNTQNSKNTPKINADISKSLIQLIYDDIKEKOKSLNSLVHGVNVPVSKDVL 233  
 QY 86 LKQNSLTPVERKHLFI-----WTGGQINDT-AINYINQWQVNSDYTVKFPVDSN-AP 136  
 DB 234 ICNDDLSKIKKIFMTDGPVFLFGQDGMGTVENIKLDNRNK-----ENNLISY 286  
 QY 137 LINTLKKTVESATNNTLESRENLENDPEFDYKFKRMEIYDKQHFIDYKSOIEE 196  
 DB 287 SINYKQVQVNNDDKKDKENINEVRD-----QKNY--YKKN--- 326  
 QY 197 NPEFIIDNIITLSNEYSKOLEALNKYIEESLNKITANNNGNDIRNLEKFADEDLVR-- 254

DB 327 -----ENINNIYDDDEK-----EDIQNKGVYNNDDI-----DEQIRKX 364  
 QY 255 -----YNQELVERWNLAAASDILRI--SMLKEDG 281  
 DB 365 MARKKYIESIPKTPKGFCLMRPVDIIDIISNTNTEMLR-----ISETLKVHENFKQHL 418  
 QY 282 GYLVDDILPGIQDPLFKSINKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKFNMDLBE 341  
 DB 419 NVLDENNSTPVVNNLLKNIN-----YK-----KNDDLIEGG 450  
 QY 342 VQSPF-----ESALLSSKSKSE-----IFLPLDDIKVSPLEVKIAPANNVIN 384  
 DB 451 EKKSFINLINVDSCYSSNSRLNENDENIERGKINNFITNDEKSIINN-----YNNNNNN 506  
 QY 385 QALISLSDSYCSDLVINQIKNRYKILNDNLNPSINEGDTFTMTKIFSDKL--ASISNE 442  
 DB 507 NNNNDNNNDNDVIIENKNNNNIYDNKN-----VECSSEKINDNGISNKN 554  
 QY 443 NMFMIKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD-----LLMFKDNSTNIHLLEPE 497  
 DB 555 --INILEPNLDT-----SNIFLEKDEYKVYVYVKNKEEIRIPLFKKEIKEI----- 599  
 QY 498 LRNPEFPKTKISQLTQEITSW-----SFNQARAKSQFEBYKGY--- 538  
 DB 600 ---FEKLPKLQYQILQDIKEWYTDNRKAIKSKDDMDVFSQVLETYVRMIKTDFFE 656  
 QY 539 -----FEGALGEDDNLQPAONT-----VLDKDYVSKILLSSMKTNRKEIHYIV 582  
 DB 657 KLIKMAENIOSVEGEL--LINKOLSKNTDINIKDYNVLOKK---KSKKKKKFLNDIL 711  
 QY 583 QLQGDKISYEASCNLFSDPYSSILYQKIEGSETAYYYVADABIKEDIKVRIYQVQISN 642  
 DB 712 N-----TYNFTTESKQDL-----YVKGESKEDIKNQIDFVTQE 746  
 QY 643 -KRNIKLTFIGHGSEFNTDFANLDVLSLSEIEITILNAKADISPKYIEINLGLCNMF 701  
 DB 747 CYRNNDIIRDTHDKS---DIFQNIKIDNNKKYIYNLELEOEINEKK-----NYN 794  
 QY 702 SYSIYAEETYFGKLLKIKDRV-SELMPISIQ---DSITVSANOYEVRI---NEEGKRE 753  
 DB 795 KNNDSNNTF---FLKIENEFKDLDDDDSQIFGDSLLADIKEYNYTADNLDNNENK 850  
 QY 754 ILDSHGKWKKEESIIDISKE---YISFNPKNK--IIVKSKYLHELSTLQIRNNA 808  
 DB 851 LYEDGENFITNEPITNEVEEKNNIYISDEQYNEEDIIFKDK-----IKERKND 903  
 QY 809 NSSDIDLEKKVMTCEINVASNDIROIVEGRIEAKNLTSD-----SINYIKNEPKLIES 864  
 DB 904 TSSD-DFE-----NCSVQ-----EKIYVNEKIEEYNNKNDKSSSSSIILEEIKYKE 951  
 QY 865 LSDSLY-----DLKHQGLDDSHF----- 883  
 DB 952 KKDELVSPNLCVLDDEFESNDLENNYISVSSDDMKTNVKNNTIGVKNKVDKTNVEYD 1011  
 QY 884 -----ISFEDISKTEGFRIRFINKETGNSIF---IETEKIFSEYATHISK--- 927  
 DB 1012 KKGDDGVIIEFDESHKLEES---KF---DNNNIYDNDDELEKNLSDYISDVKNHVN 1065  
 QY 928 EISNIKOTIFDNGVGLVKYVNLDAAEVNTLNSAFFI-----QSLIEYNTTKESLSNLSV 983  
 DB 1066 NIYNIERGEDERENEFVENKIQSTESHKSNE-----FICTENKSLRKQVMSKEDISNRI 1120  
 QY 984 AMKVQVYLAQLSTGLNTIT-----DASKVELVSTALDETIDL--PTLSGLPLII 1032  
 DB 1121 -----LAKSDIDNNLSKQNYPEILLDKQVMDNFQNIENQNDKLEKDLKDEG-AVF 1170  
 QY 1033 ATIDGVSLGAIAKELSETNDPLLRQETAEKIGIMAVNLTAASTAIVTSALGIASGFSIL 1092  
 DB 1171 EYLEDNKIISYIKETNKENNELIKYKLLKKNNEIIN----- 1208  
 QY 1093 LVPLAGISAGIPSLVNNNELIILQDKATKVIDYF--KHI--SLABTEGAPTLDDK-----II 1144

```

Db 1209 -----DEMNDIKL-----LLNFFGIPYIQSPCEAEACSYLNNKNYCDALI 1250
Qy 1145 MPQDDVLVSEIDFNNSITLGRKEIWAEGSGHTLTDDIDHFFSSPSITYRKPWLSIYD 1204
Db 1251 SDDSDLVF-----SGKTV---IKNFFN-----KKKTVEVYE 1279
Qy 1205 VLNIKKEIDFSKDLMLPNAPNRVFGYEMGWTGPRSLDNDGTKLLDRIDRHYEGQFYW 1264
Db 1280 KKAIBEKGLYQELINI---SLGCGD--YTIGVHGIG----- 1313
Qy 1265 RYFAFTADAL-ITKLPRYEDTNVRINLGNTRSFIVPVITTEQIRNLSYSFYSGGSY 1323
Db 1314 -----IVNALEIIFKAPNFEDLKILKDIVSNPFRKIDKNYNEEQ-----FLNTHKNY 1363
Qy 1324 SLS-LSPYMW-----NIDLNLVENDTWIDVNVKNI--TIESDEIQGELI 1368
Db 1364 KLNWIFPNNPPREVVKPKYKVCCTDIKKFEMHVPDIKSIYKFLHKTNISE----- 1416
Qy 1369 ENILSKLNIEDNKIILNNHTINFYGDINESNRFISLTSILEDINIIIEIDLVSYSKIL 1428
Db 1417 EKVLNVLN-----PILQYNN-----VRTYQSKIEDFPFPLE----- 1449
Qy 1429 LSGNCMKLIENSDDIOQKIDHIGFNGEHOK 1458
Db 1450 -----KKRKTVDLIDHIRANNKQKR 1470

```

Search completed: November 5, 2003, 19:29:41  
 Job time : 64.25 secs